December 9, 2017

Dear Reviewer and AMT Editorial Office,

Thank you for your detailed reading of our manuscript and the constructive suggestions you have offered for its improvement. Your review and analysis have been exceedingly helpful. We hope that we have understood and incorporated your critiques, and, in accordance with them, improved the analysis and the presentation of our work.

Please see below a detailed response to each of your suggestions and a description of how we improved our study to address your concerns. Reviewer comments are in black and our responses are in blue. Line numbers in our responses below refer to the mark-up draft of the new manuscript.

In addition to the changes in response to the reviewer's comments and suggestions, several additional changes have been made to the manuscript. These include: 1) a small correction to the flow rates for the field test data that has been discovered to be necessary since the initial submission; 2) a slight change to the method in which background methane is calculated that leads to fewer sample losses due to data gaps; and 3) resolution of a recently identified issue with beam length scaling, which has led to very small adjustments to the raw concentration data. None of these changes have altered the outcome or scientific message of the paper.

The authors are very grateful for your time and energy.

Sincerely,

Dr. Caroline Alden and co-authors

Below are detailed responses to the suggestions and comments made by the reviewers. The reviewer comments are in black and our responses are in blue.

Anonymous Referee #1:

General comments
The paper proposes a special observing system and statistical analysis to continuously detect and monitor methane leaks from gas production sites. The suggested method combines line-averaging concentration measurements with an atmospheric transport model and a novel statistical method to derive the spatially dependent methane concentration. In this way, the paper presents innovative concepts and fits within the scope of AMT.

The introduction gives a clear and concise motivation for the study, however related line-averaging measurement techniques and their advantages and disadvantages are not sufficiently addressed (e.g. Open-Path Tunable Diode Laser OP-TDL and Open- Path Fourier transform infrared spectroscopy OP-FTIR). Additionally, the non-zero minimum bootstrap (NZMB)

method is introduced without explaining the motivation behind this development in comparison to other possible methods of inverse data analysis. An extended description of the state of the art is necessary to evaluate and classify the new measurement and analysis method.
The atmospheric transport model is one of the central methods of the data analysis. The authors apply a Gaussian plume model assuming a constant methane source through the time for the synthetic data tests and for the field data under similar conditions. This approach and the application of a Gaussian plume model for the purpose of continuous methane leak detection in a real environment are extremely questionable. The Gaussian plume model for atmospheric dispersion is assuming steady-state air concentrations and mass fluxes. There are many references that show the applicability of the Gaussian model: it is primarily used to calculate seasonal or annual statistical values of air concentrations near the ground. It is recommended to use more sophisticated and realistic transport models to ensure a general applicability of the proposed method in the future. The measurement and analysis concept has a high potential of applicability if weak points of single methods will be eliminated, especially regarding the atmospheric transport model.

We thank Referee 1 for the time that they have taken in carefully reviewing this work and for their comments. In our response to the reviewer's specific comments below, we address the points raised in the paragraph above: the need for discussion of other open-path measurement techniques, and the need for discussion of the appropriateness of a steady-state transport model (namely the Gaussian plume model) for this work.

Additionally, the reviewer emphasizes a need for additional discussion of how the NZMB method fits into the context of existing inverse data analysis techniques. We have added new discussion of the state of existing methods for source location to the Introduction Section (page 3, line 28-33). Bootstrapping techniques and least-squares fitting techniques have been used for many applications in the sciences. However, there is not, to our knowledge, a field of inverse analysis devoted to minimization of false-positive sources. In response to the reviewer's suggestion, we have added introductory motivation for the development of the NZMB method to the beginning of Section 2.4 (page 7, lines 11-13). In the text we emphasize references to the traditional methods (least-squares fitting techniques and bootstrap analysis) that are combined here to produce a novel statistical methodology.

Specific comments

1. Introduction: It is highly recommended to give proper credit to related work regarding the measurement technique (line-averaging measurement methods to derive gas concentrations, e.g. OP-FTIR) and the statistical analysis methods.

Mentions and citations of other open-path laser techniques have been added to the Introduction in paragraph 3 (page 3, lines 3-5). As the focus of this paper is not the measurement technique itself, but rather the statistical methods to use line-integrated measurements as a tool for source attribution, we have kept this added discussion relatively brief while taking care to reference important previous work that does focus on the measurement technique (and which covers previous line-averaging methods in greater depth).

See response to the above comment for discussion of reference to previous work regarding the statistical analysis methods.

2. p. 2, l. 8: The 'high global warming potential' should be quantified (in comparison to carbon dioxide) to further explain the strong motivation for this study.

The GWP of CH_4 has been added to the sentence referenced by the reviewer in Section 1 (page 2, lines 8-10).

3. p. 2, l. 9: What is the reference value for the threshold leak rate in percent (3.2%)?

The reference for the threshold leak rate cited is Alvarez et al. (2012). We have added clarification and a second reference to that publication in the sentence referenced by the reviewer (page 2, line 10), in order to make it clear which of the two references at the end of that sentences pertains to the cited rate of 3.2%.

4. p. 2, l. 18: 'Cold temperatures' is a rather colloquial expression, better: lower temperature values.

The sentence cited by the reviewer (Section 1, page 2, line 20) has been edited to reflect this improvement.

5. p. 3, l. 1: What does it mean 'agreement of measurements under different conditions'? Which range of air temperature/pressure values? Which range of atmospheric stability and turbulence conditions?

The paragraph referred to by the reviewer (Section 1, paragraph 3) has been edited for clarity in this regard (see page 3, lines 5-14). In particular, we have included reference to a 2017 study by Waxman et al., who found that measurements made by two independent dual frequency comb spectrometers measuring the same air masses over a period of 2 weeks agreed under a range of ambient temperatures (4.6 – 28.9 ºC outdoors and 17 – 25 ºC indoors; the instruments were in a room that was open to the outside), and a range in relative humidity of 10-90%. A range of pressure, stability and turbulence conditions were also experienced. Furthermore, we have edited the text to clarify the principle reason why high measurement reproducibility should extend beyond ambient conditions to more extreme conditions, given that the retrieval is dependent on the quality of the absorption models and not on ambient conditions.

6. p. 3, l. 2: What does it mean 'long periods of time'? How long is it?

While the instrument does not require calibration and measurements can be directly cross-compared over periods of weeks to months (and likely years), placing an exact quantity on this length of time is involved and would require extensive additional study and experimentation. We therefore chose to simplify this statement instead. We have removed the phrase "long periods of time" and replaced it with "different conditions and times" (page 3, lines 8-9). To offer additional clarity on this point, we have also added the sentence: "Previous work also demonstrates that this method of atmospheric trace gas measurement does not require regular or

traditional calibration (Coburn et al., n.d.; Rieker et al., 2014; Truong et al., 2016; Waxman et al., 2017)" (page 3, lines 14-16).

7. p. 3, l. 4: What is the 'measurement uncertainty'? Give a short explanation of this quantity: Is it a statistical value (estimated from the standard deviation) or the technical uncertainty (depending on the device) or the analysis uncertainty or. . .?

The reviewer makes an important point that, as written, it was not clear which metric for uncertainty was being described. We have edited the text in question to clarify that the spectrometer demonstrates high measurement precision (page 3, line 17).

8. p. 3, l. 4: What does it mean 'long pathlengths'? Please be more precise or give an example or a range.

The text has been updated to reflect the length of the spectrometer paths (1 km, one-way) (page 3, lines 17-18).

9. p. 3, l. 4: Are the data and results of Coburn et al. freely available and reproducible in the meantime?

The Coburn et al. results and data are freely available and reproducible. The Coburn et al. manuscript is posted on the Cornell University Library's open access e-prints service, arXiv.org. The manuscript can be found at the following link: https://arxiv.org/abs/1711.08067.

10. p. 4, l. 5: Is the applicability of the method influenced by the special environmental conditions (e.g., wind speed and direction, atmospheric stratification) of this day in January, 2017?

The choice of January 26, 2017 was independent of wind conditions and instead depended on the preparedness of the team and instrumentation for field deployment. We have edited the text to emphasize that there were no special environmental conditions on this day (page 4, lines 22-24). Examination of long-term (1 year) mean meteorological conditions at a nearby weather station (KCOLONGM30) demonstrates that the wind speed and wind direction on this day were close to average.

11. p. 4, Gaussian plume atmospheric transport model: The application of a Gaussian plume model for the purpose of continuous methane leak detection in a real environment is extremely questionable because the model assumes steady-state air concentrations and mass fluxes. The Gaussian model maybe applied under the limited conditions of the synthetic and the real-world experiments described by the authors. However, the general applicability of the proposed measurement and analysis method is limited due to the atmospheric transport model. The cited paper of Leuning et al. (2008) contains one example of a more realistic dispersion model, the Lagrangian stochastic (LS) dispersion analysis. LS analysis can be used when the assumptions of constant and homogeneous turbulence (e.g., eddy diffusivity) are not satisfied. In contrast to the Gaussian plume model, LS models incorporate wind shear. It is highly recommended to use such

a kind of transport model to allow a general and continuous applicability of the proposed measurement and analysis method under different environmental conditions.

We agree with the reviewer that the simplicity of the Gaussian plume model with respect to assumptions of constant and homogeneous turbulence may not be as useful for field tests as, for example, a Lagrangian stochastic dispersion approach. We have added discussion of this point to Section 2.1 (page 5, lines 17-22). We have also added a discussion of this subject to the conclusions section of the paper, Section 6 (page 20, lines 20-31), which addresses these caveats to using the Gaussian plume model for representations of atmospheric dispersion.

With respect to the use of a steady-state model such as the Gaussian plume model for the present application, we are grateful for the feedback on the clarity of this point. For cases in which the time of transport from the source to the "receptor" (measurement location) is comparable to the time over which the data is averaged, a steady-state model such as the Gaussian plume model is appropriate (Gifford, 1976). In our case, the measurement averaging time is 120 seconds. In the field test cases, the mean distance between all beam locations (discretized into 100 equal-length sections) and the leaks they are monitoring, divided by the mean wind speed on 1/26/2017 of 2.1 m s^{-1}, equals a mean travel time of 100 seconds. The travel time from beam to receptor is therefore comparable to the averaging time for measurements, satisfying the requirements necessary for the Gaussian plume model assumptions of steady-state to be valid. Note that for the synthetic data cases there is no time dimension, so the use of a steady-state model is not an issue. To help better convey this point, we have added the parenthetical clarification that the plume model characterizes the steady-state in the first paragraph of Section 2.1 (page 5, line 14), and have added further clarification on page 5, lines 26-31.

For the experiments here, the assumption of constant mass fluxes is also valid, as the controlled releases of methane were constant. Future studies of intermittent leaks will need to assess transport as a function of the timescale of intermittency.

12. p. 5, l. 22: What does the term '(c/x)modelled' describes, the relationship between the point source emission and the line-averaged concentration (=spectrometer beam)? Please explain the relationship between the different scales: point source (emission) – concentration and line-averaged measurements of concentration together with the atmospheric transport model.

Section 2.1 (page 6, lines 8-15) has been edited for clarity on this point. Further, we have added a parenthetical note referring readers to Section 2.5.3 for a detailed explanation of the scaling between point source emission, to point source concentration, to line-averaged concentration (page 6, lines 14-15).

13. p. 5, section 2.2: Which spatial and temporal resolution for the derived gas concentration can be achieved using this method? How much is the environmental influence (e.g., background air temperature)?

The temporal resolution for the derived gas concentration depends on averaging time, as is shown by the Allan deviation in Figure 9 and described in Section 4.1. As averaging time increases, measurement precision increases, until such time that atmospheric variability begins to

erode measurement repeatability. We have added a sentence describing this characteristic of the DCS measurement to Section 2.2, along with a reference to Section 4.1, where it is discussed in more detail (page 6, lines 23-25).

The spatial resolution scales with beam length (which is easily adjustable by moving retroreflectors closer to or further away from the spectrometer), and beam width (which scales with telescope diameter). A discussion of this concept has been added to Section 2.2 (page 6, lines 25-27).

A discussion of the influence of environmental conditions has been added to Section 1 (page 3, lines 9-14). The addition of this text is described in greater detail in the author response to "Specific Comment 5", above.

14. p. 6, l. 4: Which fluxes? Please describe values more specific.

The paragraph referenced in Section 2.3 (page 6, line 29 – page 7, line 2) has been edited to reflect the requested improvements to specificity. Additionally, clarification has been added to Section 2.1 (page 6, lines 8-15), which is where the vector mentioned by the reviewer is described (that referenced in Section 2.3). A reference to that description has also been added to Section 2.3 (page 6, lines 31-32).

15. p. 6, l. 7: Is the system rather overdetermined or underdetermined for your examples? See also l. 19. The system is overdetermined including uncertainties/errors into the system, really?

The sentence referenced in section 2.4 has been edited to clarify the definition of overdetermined in this context: that is, that n is greater than m (page 7, line 22). Multiple measurements along each beam are used in each fit, so that m is never less than n.

The following, additional changes to the manuscript have been made to further clarify the dimensions of n and m:

In Section 2.1 (page 6, lines 10-11), the sentence was added: "In the synthetic tests and field tests described here, multiple measurements are made along each beam, such that n is always greater than m."

In Section 2.5.1, the value of m for each test is now explicitly defined (m = 20) (page 9, lines 17-18).

In Section 2.5.3 (page 10, lines 29-31), the following sentences were added: "In the synthetic tests, the dimensions of n (e.g., the length of the atmospheric concentration vector, $\mathbf{c}$) vary along with the number of beams per spectrometer-detector system and the number of meteorological conditions. In the configuration of 4 beams, for example, n = 216 * 4, because each distinct meteorological condition is applied to each beam. In the 8 beam configuration n = 216 * 8, in the 16 beam configuration n = 216 * 16, and so on."

A reference to the above section has also been added to Section 2.5.4 (page 11, line 21).

In Section 2.6.2 (page 12, lines 21-23), the dimensions of m are now specified for each case (in both cases, $m = 1$).

In Section 2.6.5 (page 13, lines 23-30), a new paragraph has been added that explicitly names the values of n for both field test cases: that for source location 1 and that for source location 2. For source location 1, all downwind measurements along beams (retros) 1 and 2 are used in the least squares fit, such that $n = 63$. For source location 2, all downwind measurements along beams (retros) 2 and 3 are used, such that $n = 30$.

16. p. 6, l. 17 and p. 7, l. 16: How did you estimate and quantify the several kinds of uncertainty? Did you only use the standard deviation as a value for the statistical uncertainty? It is highly recommendable to calculate the total uncertainty of the combined measurement and analysis method (uncertainty due to devices, measurement methods, transport model, inversion model. . .).

The reviewer brings up an important need for clarification on the subject uncertainty.

On page 6, line 17 (of the original manuscript), the reviewer points to the role of model-data mismatch uncertainty in the NZMB method. We agree that the original manuscript did not adequately explain that this source of uncertainty is *not* estimated by the user in the traditional "bottom-up" sense. That is, in many inversion frameworks, it is necessary to estimate (or in the reviewer's words, "calculate") "the total uncertainty of the combined measurement and analysis method". In the NZMB method, there is no role for "bottom-up" uncertainty assessments. Instead, the empirical fit to the data is used as the uncertainty estimate. This method inherently encompasses all of the sources cited by the reviewer (measurement device and technique, transport, representation errors, etc.) but uses the fit to the data as an estimate of such rather than a "bottom-up" estimation relying on a priori knowledge of instrumental, transport and modeling constraints. In order to address the helpful suggestion by the reviewer for clarity on this point, we have added clarifying language to Section 2.4 (page 7, lines 24-27) of the revised manuscript.

On Page 7, line 16 (and we assume also lines 17-18) (of the original manuscript) the reviewer points to the method for assessing flux uncertainty in the NZMB method. We agree that here, also, a better justification for this method of uncertainty quantification is needed. Here again, we use the empirically derived distribution of the fit to the data to assess uncertainty in the flux strength, which, as the reviewer points out, is largely a function of instrumental measurement uncertainties, transport model uncertainties, and model uncertainties. The law of large numbers justifies that when the number of bootstrap operations is large, the distribution of the bootstrapped leak strength approaches the probability distribution of the leak strength. We have added text that clarifies this point and outlines the statistical justification for this method of assessing flux uncertainty to Section 2.4 (page 8, line 33 – page 9, line 3).

17. p. 8, l. 4: What does it mean 'idealized scenario', homogeneous wind fields (wind speed and direction)?

This sentence has been revised for clarity (Section 2.5.2, page 9, lines 24-26). The text now explains why a sampling of different wind conditions on each beam is an ideal scenario for generating a large population of independent measurements.

18. p. 8, l. 7: How long is the sampling time of each beam?

In the synthetic data tests, there is no time dimension. In the field-data tests, each beam is sampled for 2 minutes.

19. p. 9, l. 2: Which influence of atmospheric stability can be expected?

The influence of parameterizations of atmospheric stability on the results of the synthetic data tests were not investigated. This is because in these tests, following conventions of most observing system simulation experiments, transport is treated as a perfectly-known characteristic of the experiment. That is, changes in the stability parameter used in the calculations for the synthetic data tests would not be expected to have an impact on the results of the study, because in these experiments the same transport characteristics that are used to create the synthetic data are also used to recover the unknown fluxes.

20. p. 9, section 2.5.4: An enhanced estimation of all parts of total uncertainty is necessary to evaluate the applied range of model-data mismatch values.

The reviewer points out that in Section 2.5.4, model-data mismatch is described as including sources of uncertainty for the simulation of atmospheric CH_4 concentrations, which includes measurement uncertainty, transport uncertainty, and representation error. In some cases, uncertainty in the characterization of background concentrations is also part of this term. However, the evaluation of the applied range of model-data mismatch values (Sect. 4.4) previously only considered the measureable quantities of that uncertainty. We find it justifiable to report only those measures of model-data mismatch that are directly measureable in our field study. However, to satisfy the reviewer's request, we have added an additional analysis that calculates a simple approximation of the transport uncertainty for the field test that is used for evaluation. Details of the simple transport uncertainty calculation are found in the Supplemental Information, and a description of the additional uncertainty this might be expected to add, along with clarifications to the existing text, have been included in Section 4.4 (page 18, lines 5-15).

21. p. 10, l. 9: Please give an example for 'long periods of time'.

The sentence in Section 2.6.1 has been edited for clarity (page 12, line 7). The implementation of the described system characteristics enable it to run autonomously for any period of time.

22. p. 11, l. 1: It can be expected that a local wind circulation is developing at Table Mountain. The assumption of homogeneity and stationarity of turbulent fields is highly questionable under these conditions (Gaussian plume model for atmospheric transport). It has to be checked that the assumptions of the Gaussian plume model are valid for the investigated location and time period.

The reviewer points out that meteorological conditions are not homogeneous or stationary through space. Measuring the entire wind field would, of course, be ideal for modeling of the environment, but the number of sensors required would not be practical. Following previous work (e.g., Hirst et al., 2004), we generalize the wind field from measurements made at a nearby location with a sonic anemometer. We have added an acknowledgement of this generalization, and the possibility that the anemometer measurements do not perfectly represent the conditions influencing the plume, to Section 2.6.4 (page 13, lines 4-7). Additionally, an estimate of transport uncertainty that includes uncertainty in the position of the wind sensor (which is assessed by examining a timeseries of data taken simultaneously by two anemometers located at different locations on the Table Mountain field site), has been added to Section 4.4 (page 18, lines 5-15) and in greater detail to the Supporting Information.

The reviewer also discusses the applicability of the Gaussian plume model for turbulent fields that are not homogeneous or stationary. The plume model endeavors to represent the mean state of fluid flow in the atmosphere, averaged over time and space, given conditions of relatively constant wind direction and speed through time and space (absolute constancy of conditions are, of course, not realistic). By using atmospheric measurements that are averaged over time and horizontally through the plume, and at a flat terrain location with no physical or topographic obstructions to flow, we find that the application of the plume model meets the basic assumptions of validity. Further, we use a short window of averaging time (2 minutes) to ensure that wind direction and wind speed conditions do not change. Finally, the Gaussian plume model requires conditions of a constant, mass-conserving pollutant source, both of which are met in these tests. In this sense, the assumptions of the Gaussian plume model are valid for the investigated location and time period. The authors agree that more sophisticated representations of atmospheric flow could be helpful for future studies focusing on source estimation in field settings. In Sections 2.1 (page 5, lines 17-20) and 6 (page 20, lines 20-31) we have added new discussion on this topic. In particular, we highlight that, while more sophisticated models are better adapted for realistic representation of atmospheric flow, there is value in applying a "Jane Doe" model that is accessible and well-known by the general scientific community. It is assumed that the use of better models would likely lead to better resulting flux estimation with the field data presented here. Using a simple model provides a baseline for performance, however, and demonstrates, at the most basic level, the potential of the proposed methodology for methane leak detection.

23. p. 11, l. 5: Can you give a reference for the applied threshold of 0.8 m/s?

We rely on the standard assumptions of Gaussian plume model application that the reliability of the method begins to decline below approximately 1 m s^{-1} (e.g., De Visscher, 2013). We have added clarification of this choice and a reference to section 2.6.4 (page 13, lines 8-9).

24. p. 14, l. 2: The (short) averaging time of 1-2 min disagrees probably with the assumption of stationarity of the Gaussian plume model. It is questionable if the assumptions of atmospheric transport model are satisfied. At least, an enhanced contribution of uncertainty should be included in the analysis (and conclusions).

We refer the reviewer to the edits to the manuscript in response to "Specific Comment" number 11, above, in response to the comments here regarding the applicability of a steady-state model over a 120-second averaging time, and the assumptions of the atmospheric transport model used. In summary of that point, it has been shown that when measurement averaging time is comparable to transport time between source and measurement point, a steady-state model such as the Gaussian plume model is appropriate (Gifford, 1976). For added clarity on this subject, we have edited the text in Section 2.1 (page 5, lines 27-30).

The reviewer makes a helpful suggestion regarding clarification of whether or not contributions of uncertainty from the choice of or assumptions associated with the transport model are included in the analysis. With respect to inclusion of enhanced contribution of uncertainty (due to choice of transport model) in the analysis:

In the synthetic data tests, the levels of "noise" added to the data are not meant to represent quantifiable combinations of the various sources of model-data mismatch uncertainty. That is, added noise of 1 ppb is not, for example, broken down into the components of uncertainty contributing to that overall value. In that sense, contributions of transport uncertainty are indeed included in the synthetic tests.

In the field data tests, so-called "bottom-up" estimations of observation uncertainty are not part of the analysis, but uncertainty due to transport modeling is, again, an important and included part of the analysis. The NZMB method first uses an NNLS approach to fitting a flux value to the observations, for which estimation of "bottom-up" observation (i.e., model-data mismatch) uncertainty is not necessary. In the second step, the NZMB method uses the actual residuals of the fit to the data as a measure of uncertainty in that term. This means that contributions of transport uncertainty are included in the observation (i.e., model-data mismatch) uncertainty.

We address the need for clarification of this point with added text to Section 2.4 (page 7, lines 24-27), in response to "Specific Comment" number 16, above.

In response to the reviewer's suggestion that a discussion of transport model contributions to uncertainty be added to the Conclusions, we point to added text in Section 6 of the paper that addresses this point.

25. p. 16, section 6: The critical discussion of the used atmospheric transport model is missing here. If a simple assumption (Gaussian plume model) is used for real (complex) environmental conditions, a comprehensive analysis of uncertainties is needed to evaluate the potential and the general applicability of the method.

We have added substantial discussion to the paragraph cited here at the reviewer's request (page 20, lines 20-31). The simplicity of the Gaussian plume model in achieving the correct identification of leaks in a field test serves here to demonstrate the viability of the measurement system and statistical technique for leak detection. Included in the added text is discussion of the need for a future analysis of transport models to identify which would be most suitable for the identification of leaks using the DCS system and NZMB method. This comprehensive analysis

would be best suited for a future study of that problem, as to undertake such an analysis would not be directly related to the central themes and goals of the work in discussion here.

26. Fig. 10: Are the concentration data referred to the length of the beam? Please provide the wind direction according to meteorological conventions: 0 deg = north, 180 deg = south, 90 deg = east, 270 deg = west). Which averaging time was used to provide the data?

The concentration data are corrected for beam length. The data shown therefore depict line-averaged (or line-integrated) concentrations, as described in the text.

The caption for Figure 10 has been edited to specify wind direction according to meteorological conventions, as recommended by the reviewer.

The averaging time used to provide the data is 2-minutes. This has been clarified in the caption for Figure 10.

Technical corrections
1. p. 3, l. 4: 'Coburn' – 'et al.' is missing.

This typo has been corrected.

2. p. 5, l. 28: 'Truong et al.; Waxman et al.' – The year is missing.

This typo has been corrected.

3. p. 5, l. 32: 'Coburn' – 'et al.' is missing.

This typo has been corrected.

4. p. 5, l. 33: 'tempus lorem' – misprint?

This typo has been corrected.

5. p. 8, l. 8: 'that that' – delete one 'that' 6. Table 2: Source Location 1 and NZMB solution: unit must be kg/s

This typo has been corrected.

Anonymous Referee #2:

In atmospheric "inverse dispersion" problems, we attempt to deduce something about gas transfer to or from the atmosphere by (i) sampling the gas concentration field, (ii) simultaneously gathering sufficient meteorological data to characterize atmospheric transport (mean wind and turbulence), then (iii) invoking an atmospheric dispersion model to make inferences about the source(s). Within that overall scheme there are many variants. Such problems arise and can be solved on every meteorological scale of atmospheric (or, for that matter, oceanic) motion. Sometimes one knows the exact location of the source(s) and the object is their quantification; sometimes deducing source location (in space and or time) is the most important aim; and sometimes one seeks to deduce source space-time location *and* strength.

This paper is highly relevant to the problem of determining by inverse dispersion the spatial location(s) of point- or near-point sources in the context of gas leak detection (specifically methane) from industrial sites, and an element that is stressed is the avoidance of false positives (for potential leak locations) while not failing to find actual (non-zero) source locations. The authors briefly describe what appears to be a highly capable and useful instrument for their purpose (though describing the instrument is not the key aspect of this paper), but mostly focus on their statistical strategy for optimizing the useful information that can be extracted from their "measurements" (the quotes, because the paper invokes measurements both synthetic and real). I have not been able to fully comprehend that statistical strategy, and I feel the recipe for it can and should be clearer. Some readers may wonder why a Bayesian framework has not been adopted, as perhaps the most rational way to make use of prior information and account for uncertainties.

We thank Referee 2 for the time that they have taken in carefully reviewing this work and for their helpful comments.

We have edited the text in Section 2.4 (page 7, line 31 – page 8, line 13) to clarify the recipe for the statistical strategy. The changes to the text include expanded descriptions of the methodology and simplifications of the nomenclature.

The reviewer raises an important question regarding why a Bayesian framework was not adopted. Given the lack of clarity in the previous draft of the manuscript regarding the dimensions of m and n (which we have tried to address in the new draft, particularly in response to the Specific Comments, below), we see why the reviewer would have suggested a Bayesian approach. Given that n is greater than m in the flux estimation problems presented here, we find that our approach is suitable, compared with approaches (such as the Bayesian approach) that are typically used for ill-conditioned (under-determined) problems, or problems in which the number of data points used to inform the fluxes are equal to or fewer than the number of fluxes to be resolved.

In this manuscript, we describe a leak detection strategy with the dual goals of identifying and quantifying potential leaks while limiting the instances of false positive leak detection. The

proposed method shows promise to address both goals with accuracy.

Setting aside that aspect of the paper, to my mind the most vulnerable aspect of the authors' methodology is their use of the Gaussian plume model as their atmospheric transport paradigm, which treats the turbulent surface layer wind as if it were a regime of unsheared homogeneous turbulence. The authors do recognise that the Gaussian plume model is highly simplistic, and I can accept their argument that its use in the context of their paper is acceptable. But to drive home the importance of the choice of wind model, I must stress that the mean wind speed and the effective eddy difffusivity vary radically with height across the atmospheric surface layer (ASL), in a manner that is well described by Monin-Obukhov similarity theory. There are much higher fidelity models available, and one of these could be substituted without great penalty in terms of computational burden. It is true that at some sites, obstacles or topography may disturb the transporting wind field such that it is more complex than envisaged even by the better models (some of which are listed below), but they cannot be a worse choice than the Gaussian plume model, which is in effect a mental straitjacket.

We recognize the reviewer's concerns about our use of the Gaussian plume model, and address this issue in our response to the related comment number 1, below.

Specific Comments

1. The authors' admit that the Gaussian plume model (GPM) is highly simplistic, and I accept their argument that its use in the context of their paper is acceptable. I would disagree with their *categorical* (i.e. no exceptions) assertion that the GPM is "more suitable" if the inversion is based on line-averaged concentration data. The GPM is unnecessarily simple. There are much better analytical solutions that could and should be used to evaluate the H matrix. Whereas the GPM entirely neglects mean wind shear and treats the eddy diffusivity as constant (the atmospheric surface layer being represented as a regime of unsheared homogeneous turbulence), better models represent the ASL with a mean wind shear and a height-dependent eddy diffusivity that are consistent with the state of the ASL as parameterised by the Obukhov length and friction velocity. These solutions can be rapidly calculated; and where they provide (only) cross-wind integrated concentrations, the authors could easily introduce Gaussian *crosswind* spread (note: it is hard to improve on the assumption of Gaussian crosswind spread without using measured data on wind direction fluctuations). A better analytical (or semi-empirical) eddy diffusion model could be sought out from the following references: Philip (1959), Ermak (1977), van Ulden (1978), Nieuwstadt and van Ulden (1978), Huang (1979), Lupini et al. (1981), Wilson (1982), Kormann and Meixner (2001), Sharan & Kumar (2009), Wilson (2015).

The reviewer is correct; the sentence the reviewer refers to ("more suitable") was poorly phrased and led to the implication that the Gaussian plume model is more suitable than other models in this context, which is not the case. We emphasize here and with additional text in the manuscript (see Section 2.1 and Conclusions) that, while more sophisticated models are better adapted for realistic representation of atmospheric flow, there is value in applying a "Jane Doe" model that is accessible and well-known by the general scientific community. It is assumed that the use of better models would lead to better resulting flux estimation with the field data presented here. Using a simple model, however, provides a baseline for performance and demonstrates, at the most basic level, the potential of the proposed methodology for methane leak detection.

We have replaced the text referred to by the reviewer (Section 2.1, page 5, lines 15-24) with a note on the drawbacks of the plume model, and recommendations that future studies focused on field data or applications of the methodology for leak detection purposes consider employing a more sophisticated plume model such as those suggested, AERMOD or a stochastic Lagrangian particle model. Additionally, we have included additional discussion – in the Conclusions section of the manuscript (page 20, lines 20-31) – that more realistic representation of atmospheric transport could be achieved with the use of alternative models.

2. p5, line 30: Is there an easy argument that the optical detector's response is to line-averaged methane mole fraction as opposed to line averaged mass concentration [kg m^{-3}]?

The reviewer is correct on this matter; the absorption measurement is proportional to the concentration (number density) of the absorbing gas along the line of site (laser beam) rather than directly related to the mole fraction - which is calculated using the temperature (derived from the absorption spectrum) and pressure (from an external pressure monitor).

We have adjusted the text to reflect the reviewer's important clarification on this matter (page 6, line 21).

3. At p6 line 5, why "attempts to solve"?

The reviewer is correct. The NNLS algorithm does not "attempt to solve" the least squares problem; it computes value(s) of **x** that solve the least squares problem. We have edited the text highlighted by the reviewer for accuracy in this regard (page 7, line 1).

4. I don't understand why the authors contend (p6, line 19) that (in general, with m source locations and n concentrations) their "problem is overdetermined" — do they assume n > m?

The dimensions *n* and *m* were previously not clearly defined, and the reviewer has helpfully identified this problem in both this comment and comment number 7, below. In fact, the dimensions of *n* are greater than the dimensions of *m*, such that the problem is overdetermined. Multiple measurements along each beam are used in each fit, so that *m* is never less than *n*.

To clarify this important point, we have adjusted the text in the following sections:

In Section 2.1 (page 6, lines 10-11), we have added the sentence: "In the synthetic tests and field tests described here, multiple measurements are made along each beam, such that *n* is always greater than *m*."

In Section 2.4 (page 7, line 22), in the sentence identified by the reviewer in this comment, we have added the clarification: "the problem is overdetermined (that is, $n > m$)."

In Section 2.5.1 (page 9, lines 1718), the value of *m* for each test is explicitly defined ($m = 20$).

In Section 2.5.3 (page 10, lines 29-32), we have added the clarifying sentences: "In the synthetic tests, the dimensions of n (e.g., the length of the atmospheric concentration vector, $\mathbf{c}$) vary along with the number of beams per spectrometer-detector system and the number of meteorological conditions. In the configuration of 4 beams, for example, $n = 216 * 4$, because each distinct meteorological condition is applied to each beam. In the 8 beam configuration $n = 216 * 8$, in the 16 beam configuration $n = 216 * 16$, and so on." A reference to this Section has also been added to Section 2.5.4 (page 11, line 21).

In Section 2.6.2 (page 12, lines 21-23), the dimensions of m are specified for each case (in both cases, $m = 1$).

In Section 2.6.5 (page 13, lines 23-29), a new paragraph has been added that explicitly names the values of n for both field test cases: that for source location 1 and that for source location 2. For source location 1, all downwind measurements along beams (retros) 1 and 2 are used in the least squares fit, such that $n = 63$. For source location 2, all downwind measurements along beams (retros) 2 and 3 are used, such that $n = 30$.

5. Why (p6 lines 22-23) should it be the case that (or why is it a safe assumption that) "model-data mismatch uncertainty has an un-biased Gaussian distribution"?

The reviewer is correct in pointing out that it is not a safe assumption to suppose that there is no bias in the distribution of the model-data mismatch uncertainty. We have amended the text to reflect this point (Section 2.4, page 7, lines 28-29). For the purposes of the methods and analysis here, we keep the un-biased model-data mismatch because we believe it will take a separate and extensive study to examine realistic distributions of uncertainty for the test cases posed. This assumption follows on previous observing system simulation experiments and other work focused on emissions detection. For example, from Crenna et al. (2008): "Model error also arises unavoidably because numerical models are based on idealized relationships that are only an approximation of the real world. Such idealizations presumably introduce systematic errors in a_{ij}, but it is outside the scope of this paper to attempt to address such errors."

6. I find the derivation (p6 line 31 to p7 line 2) hard to follow, yet I suspect it has to be very simple. For instance we have the three symbols ε_{Ri}, ε_R and ε_{bi}: is the last just the first, in alternative guise?

The reviewer is correct that ε_{Ri} and $\boldsymbol{\varepsilon}_R$ are the same (the latter is simply in vector notation), and ε_{bi} is similar: it represents a bootstrapped sampling of $\boldsymbol{\varepsilon}_R$. That is, ε_{bi} is a permutation or resampling of ε_{Ri}. In the text, we have removed the subscript R and replaced ε with e, which is a more correct notation, so that the two vectors are now simply $\mathbf{e}$ and $\mathbf{e}_b$, to make it clearer to the reader that the latter is a bootstrap sampling of the former, not a new value altogether (page 7, line 31 – page 8, line 13).

7. If I have understood correctly, the pool of residual values is a set containing only n members, where n is the number of concentration measurements (n = 3 for the field test). Then, for each detector (i = 1...n) one randomly draws 1000 samples from that set, with replacement, thereby

obtaining a set of 1000 alternative model predictions y_{bi} for (each) source location. The logic for this is not very sound, it seems to me, because all observations are given equal status, irrespectively of their distance from the source(s). In real world cases there could be order-of-magnitude differences in the measured mean concentrations — and indeed in concentration variance and higher moments that, although irrelevant here, surely relate to the trustworthiness or representativeness of a measurement — and in the level of uncertainty in the modelling.

Firstly, we thank the reviewer for pointing out that the values of *n* (and indeed *m*) require clarification in the manuscript. The value of *n* is substantially larger than 3 in the field tests (*n* is 30 or more) and in the synthetic tests (*n* = 864 or more). We refer to the text changes documented in response to comment number 4 for clarity on this matter.

The reviewer brings up a second important point regarding the treatment of all observations as having "equal status". For the field tests, it becomes evident through the clarifications outlined in response to comment number 4 that the observations used in the fit (and therefore in the bootstrap) occur on the same beam, if (as is the case in the field data presented here) the wind direction is such that the downwind direction does not change throughout the course of sampling. For that reason, different observation residuals resampled in the bootstrap do have "equal status" in terms of the distance between source and receptor.

However, a corollary to the reviewer's point is that, even in the field test cases (in which the distance to the downwind beam does remain static throughout the experiment) changes throughout the measurement period in the meteorological conditions may mean that not all observations should be treated equally in the bootstrap analysis. For example, differences in the measured concentrations due to changes in wind speed, wind direction, and atmospheric stability could be expected to result in changes in the measurement representativeness and uncertainty in modeling.

To address this issue, we have adjusted our analysis to now implement a moving block bootstrap (replacing the bootstrap used in the NZMB method). The moving block bootstrap method recognizes the short-range dependence of measurements collected under similar conditions and, relatedly, the independence of measurements not collected under similar conditions. We refer the reviewer to Künsch (1989) for a full description of the moving block bootstrap technique. We calculate the autocorrelation of the residuals to determine the length of time appropriate for the moving block bootstrap, or the time "window" over which residuals are resampled for a given observation. We take two times the length of time at which residuals demonstrate autocorrelation at the 95% confidence level. The autocorrelation length is doubled to ensure that a sufficient number of observations are included in the moving window to provide a robust statistical sample size. For the measurements at source location 1, the moving block window is 78 minutes in length and for the measurements at source location 2, the moving block window is 114 minutes in length.

The same change in methodology is not applied to the synthetic tests because there is no time dimension over which to apply a moving block bootstrap. Field tests or synthetic tests using time-resolved meteorological data could use the moving block bootstrap, and the block bootstrap

could further be restricted to sampling of residuals on a beam-by-beam basis for source-receptor configurations that more closely mirror those in the synthetic tests.

Text has been added to section 2.4 (page 8, lines 1520) to describe and reflect the methodological updates described here.

8. At p7, line 13, it might be helpful to be more specific as to what "law of large numbers" means in this context. I expect it amounts to an assumption that the distribution of some mean value (or sum) is Gaussian, even if the numbers being summed do not have a Gaussian distribution (central limit theorem)?

We agree that clarification on this matter is needed, and have added text to Section 2.4 (page 8, line 32 – page 9, line 3) to that end. We rely on statistics literature, which states that, even for the dependent case, under certain regularity conditions on random variable moments, the sample mean of the random variables approaches the population mean. This holds true for the bootstrapped case as well. That is, when the number of bootstrap operations is large, and given the sample, the bootstrapped leak mean approaches the estimated leak from the sample ("law of large numbers"). The Central Limit Theorem (CLT) states that under appropriate scaling for large sample size, the test-statistic (or estimated leak rate) converges in distribution to a normal distribution. When the number of bootstrap operations is sufficiently large and n is also large, then, assuming the residuals are independent, it has been shown (Bickel and Freedman, 1981; Singh, 1981) that the bootstrapped leak mean admits a CLT. That is, the bootstrapped mean converges in distribution to a normal distribution with mean being the unknown population leak (or leak to be estimated) and scaling factor as the square root of n.

9. At p9 line 21, the authors allude to "model-data mismatch noise." It seems to me that "noise" type errors (which can be dealt with by averaging) are in practice likely to be far less serious than *systematic* errors arising from the imperfect modelling of atmospheric transport.

The reviewer is correct that the word "noise" evokes the kind of uncertainties that can be easily averaged out, whereas model-data mismatch uncertainties may not always be Gaussian. In the paragraph in question, we use the word "noise" rather than the word "uncertainty" to emphasize that we perturbed the synthetic data to create an observing system simulation experiment (OSSE). We do, however, change several instances of the word "noise" to "uncertainty", as the discussion in the first part of the paragraph does focus on real uncertainties associated with transport and measurement precision. The title of section 2.5.4 (page 11, line 1) has been change to reflect the reviewer's point, and text edits have been made throughout that section (particularly page 11, line 15-20).

A related point that the reviewer makes is that Gaussian noise is potentially less problematic for inversions than would be systematic errors or uncertainties in the ability to simulate observations. Text added to Section 2.4 (page 7, lines 28-29) acknowledges this possibility. For the purposes of the study presented here, we apply the traditional OSSE framework of perturbing the observations with Gaussian noise. Ideally, measurements would be gathered for a long enough period of time so that transport and other errors would assume a Gaussian distribution, however this is not always possible in limited-scale field studies and may not always be the case

regardless of the length of the dataset, given dispersion model uncertainties. Future studies of systematic biases in model-data mismatch and impacts on flux estimation would therefore be warranted.

10. How is gas-gas interference dealt with? What about detection of "stray" infra-red radiation emitted from the environment and/or from within the telescope?

Gas-gas interference is avoided in dual-comb spectrometry due to the large spectral bandwidth (~60 nm) and simultaneous high spectral resolution (0.002 nm). The bandwidth and resolution allow the system to fit and distinguish the individual absorption features and patterns for each gas - even when the patterns overlap.

"Stray" or background infrared light does not affect the laser signal measured on our fast photodiode detector due to the heterodyne nature of the detection. "Stray" signals do appear at low frequencies, but these signals do not affect the high frequency heterodyne beat signals between the comb teeth in dual-comb spectrometry.

We have added clarification of these points in Section 2.6.1 (page 11, line 31 – page 12, line 4) to help readers who will have similar questions.

11. Inversion of the H matrix can entail severe error if the matrix is "ill-conditioned" as a result of the relative positioning of the sources and detectors: for example if sources are aligned along the wind direction (see Crenna et al. 2008; Flesch et al. 2003).

The reviewer raises an important point and suggests several seminal papers on the topic of sensor placement for optimization of the condition number of the H matrix. In the experiments presented here, we sought to test a case study that was agnostic to sensor placement, as that topic merits a full study of its own for the particular distribution of point sources such as is posed by oil and gas operations. By randomly distributing beams among wells, assuming known background concentrations, and incrementally increasing beam density in a way that is agnostic to point source location, we hope to provide the simplest test of the NZMB method. This point is made in the study's Conclusions Section 6, but, as the reviewer points out, it needs to be emphasized. We have added text to that section of the paper citing the papers referenced by the reviewer here as examples of studies focused on the benefits (and drawbacks, if not properly considered!) that sensor placement can pose (page 20, lines 13-17).

In order to create a field experiment that closely matched the synthetic experiment condition of perfectly-known background concentrations, we relied on sensor placement such that the background concentration was measured independently from signals arising from local sources, such as Crenna et al. (2008) suggest. Similarly, sensors are placed so that a given sensor is unlikely to measure emissions from more than one source, achieving "two nearly independent single-source problems", as also recommended by Crenna et al. (2008). These beam placements serve to minimize k, and will certainly inform future studies aimed at sensor placement in complex or dense fields of wells and facilities.

Finally, as Crenna et al. (2008) points out, more measurements (particularly under different

sampling conditions) can lead to lower condition numbers, even in cases of sub-optimal sensor placement. The synthetic experiments here are designed to capitalize on this characteristic, by simulating a wide range of wind conditions. Field experiments and future field sampling of emissions from real oil and gas facilities will similarly benefit from longer measurement timeseries to maximize the number of independent tracer interactions with sensors.

To address these points, the text has been edited and appropriate references included in Section 2.5.2 (page 9, lines 24-26 and page 10, lines 1-3), Section 2.6.2 (page 12, lines 17-19), and Section 6 (page 20, lines 13-17).

12. Using the GPM entails selection of appropriate σ-curves: the choice should be documented.

The reviewer is correct – thank you for identifying this oversight. We have added a description in Section 4.2 (page 16, lines 26-29) of the stability classes and σ-curves used for the field data, and have added more detailed description of the calculation of the σ-curves in the Supplemental Information.

13. The Conclusion does not, to my mind, sufficiently recognise the potential accuracy gain from using a more sophisticated and realistic atmospheric transport model.

The co-authors agree that this issue needs to be revisited again in the concluding section of the paper. A paragraph addressing the use of the GPM (and recommendations to use more realistic simulations of transport) has been added to the final (Conclusions) Section (page 20, lines 20-31).

14. I wondered whether the authors are familiar with normal practice in regard to averaging, in micrometeorology. To make sense of surface layer winds of course we necessarily must use statistics, and it is usual to base those on an averaging interval of at least about 15 minutes. However the discussion caused me to suspect that the authors envisage using nearly-instantaneous wind measurements (or, say, 1 min averages) to deduce leak rates from concentration averages over (say) one or two minutes. This may be feasible, indeed it may be a good idea (at the least, it would speed up a search across many potential leak sites). The thing is, however, that one has to recognize that existing, documented, tested, trusted surface layer dispersion models exploit a statistically stable estimate of surface layer state, and that statistical stability demands averaging intervals much longer than a minute (or two). To adapt existing atmospheric transport modes to very sort averaging intervals one will at the least have to adjust the parameterisation of the eddy diffusivity.

The reviewer makes a valid point that it is potentially problematic to apply parameterizations of atmospheric dispersion to 2-minute datasets that were developed for longer time scales, and generally to interpret dispersion models at short timescales. The reviewer aptly identifies that it is valuable for the purposes of the application presented here to obtain information over short time frames, in particular to speed up the search across many possible leak locations. Further, this approach is necessary because background methane concentrations fluctuate with extremely high frequency (order minutes) in proximity to natural gas operations. Observational datasets

that would average across these fluctuations would likely miss the small signals that we attempt to recover here.

While this issue may remain a substantial and unresolved caveat of this methodology, we point to several characteristics of the observing framework that may help to mitigate its impact. We also perform a sensitivity test to analyze the impact of adjustments to the eddy diffusivity parameterizations.

Emissions from oil and gas are typically close to the ground surface (natural gas wells are not generally taller than several meters in height). Our sensors are typically also located near the surface. For this reason, vertical fluctuations in the plume between the source and receptor may be expected to be truncated by interaction with the ground surface, which, in theory, would mean that sensitivity to shorter time period oscillations should be damped in comparison with source-sensor configurations located higher above the ground level. Weil et al. (2012) demonstrate this idea, for example, by showing that crosswind-integrated concentration fluctuation intensity is smaller for sources closer to the surface than for elevated sources. This provides at least some assurance that shorter averaging times may be less problematic at the surface than higher above ground level.

Despite the US EPA's recommendation that hour-long averaging times be used for application of the Briggs dispersion parameterizations, the Briggs parameterizations were derived using 10-minute averaged data (Beychok, 2005). The Australian EPA recommends use over 3-minute time frames. It is possible, therefore, that the discrepancy between the time averaging used in the work presented here is not as vast as, for example, the EPA recommendations would imply. In recognition of the discrepancy between the averaging time for which the Briggs parameterizations were developed (10 minutes) and the averaging time used here (2 minutes), we calculate the expected potential error introduced in sigma-y using the standard equation for the adjustment of dispersion parameters for different averaging times:

$$\sigma_{y,2} = \sigma_{y,1} \left(\frac{t_2}{t_1}\right)^p$$

This equation can be found in Gifford (1976).

If the Briggs parameters used here were calculated using an averaging time of 10 minutes, then t_1 = 10 minutes. We use measurements averaged over 2 minutes and meteorological data averaged over 2 minutes, so that t_2 = 2 minutes. Using the typical value of 0.2 for the empirical parameter p (for t_2 < 1 hour), we anticipate a factor of 0.7 correction to the Briggs dispersion parameter σ_y.

Using the above recommended method for adjustment of σ_y, we re-calculate the results shown in Table 2 as a sensitivity test. We are not aware of an equivalent equation for the adjustment of σ_z. We find that the sensitivity of our result to the adjustment to σ_y is negligible: the value of the non-bootstrap solution for Source Location 2 changes by 0.1 E-5 kg s^{-1}. This is, perhaps, to be expected, given that the cross-wind fluctuations are largely averaged through by the laser beam when winds are orthogonal to it. We report this sensitivity and show histograms of the results from this sensitivity test in the Supplemental Information section.

	Source Location 1	Source Location 2
Controlled Leak Time On:	10:08	NA
Controlled Leak Time Off:	16:30	NA
Measured Mean Flow Rate:	3.1 E-5 ± 0.01E-5 kg s^{-1}	0.0 ± 0.0 kg s^{-1}
Non-Bootstrap Solution:	2.4 E-5 kg s^{-1}	0.5 E-5 kg s^{-1}
NZMB Solution:	2.6 E-5 ± 0.5 E-5 k/s	0.0 ± 0.0 kg s^{-1}

A final note on averaging time, and perhaps the most relevant defense of the methods we lay out here, is that the separate dispersion calculations (which are based on 2-minute averages) are aggregated together in a single least-squares fitting routine that spans much longer time frames (several hours). This means that unresolved or unresolvable errors in the vertical and horizontal dispersion over each 2-minute interval may be resolved, in the mean, over longer time periods. This approach is similar to that taken in CALPUFF, for example: when averaging times exceed 1 hour, the algorithm applies averaging of each separate 1-hour dispersion calculation (Scire et al., 2000).

We find, for the goals of the methods developed here, that the losses in model fidelity incurred by averaging over short timescales (2 minutes) are acceptable given the particular requirements of methane leak detection (resilience to rapid background variability, rapid scanning for potential leaks). This is particularly true given that 2-minute intervals are aggregated over longer time periods to obtain solutions. The promising results shown here suggest that this trade-off, of more accurate parameterizations for a useful methane leak detection solution, might be worth making.

We have included a discussion of this important topic in Section 5.2 (page 19, line 22 – page 20, line 3) of the paper.

Technical corrections

1. First paragraph of Sec. 2.6.2 uses mixed tenses (present then past).
   The tenses have now been made the same in this section.

2. Figure 3 shows only two beams for the field test, whereas the discussion of the experiment alludes to three.
   Thank you for catching this mistake. We have replaced Figure 3 with the correct version, which shows all three beams and both potential leak locations.

3. The number of panels on Figures (5,6) seems excessive — is it necessary to cover the range in MDM (model-data-mismatch) with such fine steps.
   Thanks to the reviewer for this suggestion. We have adjusted Figures 5 and 6 to only show a subset of the MDM results: 0.5 ppb, 1.0 ppb, 2.5 ppb, 5.0 ppb, and 10.0 ppb. Accordingly, the text in the paragraph beginning on page 15, lines 5-7 has been adjusted. Language was also added in Section 3.1. to guide the reader in the interpretation of these figures, as they are complex. An issue with figure numbering has also been fixed. We hope that these adjustments improve readability. We have included the full range of MDM in the Supplemental Information.

4. The structuring of this paper results in a degree of repetition.

We removed repetitive phrases, sentences or paragraphs from the following sections: 1, 2.1, 2.4, 2.5.2, 2.5.3, 3.2, 4.3

References

Crenna, B.P., Flesch, T.K., & Wilson, J.D. 2008. Influence of Source-Sensor Geome- try on Multi-Source Emission Rate Estimates. *Atmos. Environ.*, **42**, 7373–7383. DOI: doi:10.1016/j.atmosenv.2008.06.019.

Ermak, Donald L. 1977. An Analytical Model for Air Pollutant Transport and Deposition from a Point Source. *Atmos. Environ.*, **Vol. 11**, 231–237.

Flesch, T.K., Harper, L.A., Desjardins, R.L., Gao, Z., & Crenna, B.P. 2009. Multi-source emission determination using an inverse-dispersion technique. *Boundary-Layer Meteorol.*, **132**, 11–30.

Huang, C.H. 1979. A Theory of Dispersion in Turbulent Shear Flow. *Atmos. Environ.*, **13**, 453–463.

Kormann, R., & Meixner, F.X. 2001. An Analytical Footprint Model for Non-Neutral Stratification. *Boundary-Layer Meteorol.*, **99**, 207–224.

Lupini, Renzo, & Tirabassi, Tiziano. 1981. A Simple Analytic Approximation of the Ground-Level Concentration for Elevated Line Sources. *J Appl Meteorol*, **20**, 565–570.

Nieuwstadt, F. T. M., & Ulden, A. P. Van. 1978. A Numerical Study on the Vertical Dispersion of Passive Contaminants from a Continuous Source in the Atmospheric Surface Layer. *atmos. envt*, **12**, 2119–2124.

Philip, J. R. 1959. The Theory of Local Advection. *J. Meteorol.*, **16**, 535–547. (I have clarifying correspondence with JRP in a folder.).

Sharan, M., & Kumar, P. 2009. An Analytical Model for Crosswind Integrated Concentrations Released from a Continuous Source in a Finite Atmospheric Boundary Layer. *Atmos. Envi- ron.*, **43**, 2268–2277.

van Ulden, A. P. 1978. Simple Estimates for Vertical Diffusion From Sources Near the Ground. *Atmos. Environ.*, **Vol. 12**, 2125–2129.

Wilson, J.D. 1982. An Approximate Analytical Solution to the Diffusion Equation for Short-Range Dispersion from a Continuous Ground-Level Source. *Boundary-Layer Meteorol*, **23**, 85–103.

Wilson, J.D. 2015. Dispersion from an area source in the unstable surface layer: an approximate analytical solution. *Q.J.R. Meteorol. Soc.*, **141**(693), 3285–3296.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-262, 2017.

References

Alvarez, R. A., Pacala, S. W., Winebrake, J. J., Chameides, W. L. and Hamburg, S. P.: Greater focus needed on methane leakage from natural gas infrastructure, Proc. Natl. Acad. Sci., 109(17), 6435–6440, doi:10.1073/pnas.1202407109, 2012.

Beychok, M. R.: Fundamentals of Stack Gas Dispersion, 4th ed., Newport Beach, CA., 2005.

Bickel, P. J. and Freedman, D. A.: Some asymptotic theory for the bootstrap, Ann. Stat., 9(6), 1196–1217, doi:10.1214/aos/1176345637, 1981.

Coburn, S., Alden, C. B., Wright, R., Cossel, K., Baumann, E., Truong, G.-W., Giorgetta, F., Sweeney, C., Newbury, N. R., Prasad, K., Coddington, I. and Rieker, G. B.: Continuous regional trace gas source attribution using a field-deployed dual frequency comb spectrometer, submitted,

n.d.

Crenna, B. P., Flesch, T. K. and Wilson, J. D.: Influence of source – sensor geometry on multi-source emission rate estimates, Atmos. Environ., 42, 7373–7383, doi:10.1016/j.atmosenv.2008.06.019, 2008.

Gifford, F. A.: Atmospheric dispersion models for environmental pollution, in Lectures on Air Pollution and Environmental Impact Analysis, edited by D. A. Haigen, pp. 35–58, Boston, Mass., 1976.

Hirst, B., Gibson, G., Gillespie, S., Archibald, I., Podlaha, O., Skeldon, K. D., Courtial, J., Monk, S. and Padgett, M.: Oil and gas prospecting by ultra-sensitive optical gas detection with inverse gas dispersion modelling, Geophys. Res. Lett., 31, 1–4, doi:10.1029/2004GL019678, 2004.

Künsch, H. R.: The jackknife and the bootstrap for general stationary observations, Ann. Stat., 17(3), 1217–1241 [online] Available from: http://www.jstor.org/stable/2241719, 1989.

Rieker, G. B., Giorgetta, F. R., Swann, W. C., Kofler, J., Zolot, A. M., Sinclair, L. C., Baumann, E., Cromer, C., Petron, G., Sweeney, C., Tans, P. P., Coddington, I. and Newbury, N. R.: Frequency-comb-based remote sensing of greenhouse gases over kilometer air paths, Optica, 1(5), 290–298, doi:10.1364/OPTICA.1.000290, 2014.

Scire, J. S., Strimaitis, D. G. and Yamartino, R. J.: A User's Guide for the CALPUFF Dispersion Model, Concord, MA., 2000.

Singh, K.: On the Asymptotic Accuracy of Efron's Bootstrap, Ann. Stat., 9(6), 1187–1195, doi:10.1214/aos/1176345636, 1981.

Truong, G.-W., Waxman, E. M., Cossel, K. C. C., Baumann, E., Klose, A., Giorgetta, F. R., Swann, W. C., Newbury, N. R. and Coddington, I. C.: Accurate frequency referencing for fieldable dual-comb spectroscopy, Opt. Express, 24(26), 30495–30504, doi:10.1364/OE.24.030495, 2016.

De Visscher, A.: Air Dispersion Modeling: Foundations and Applications, John Wiley & Sons, New York., 2013.

Waxman, E. M., Cossel, K. C., Truong, G.-W., Giorgetta, F. R., Swann, W. C., Coburn, S., Wright, R. J., Rieker, G. B., Coddington, I. and Newbury, N. R.: Intercomparison of Open-Path Trace Gas Measurements with Two Dual Frequency Comb Spectrometers, Atmos. Meas. Tech. Discuss., 1(March), 1–26, doi:10.5194/amt-2017-62, 2017.

Weil, J. C., Sullivan, P. P., Patton, E. G. and Moeng, C. H.: Statistical Variability of Dispersion in the Convective Boundary Layer: Ensembles of Simulations and Observations, Boundary-Layer Meteorol., 145(1), 185–210, doi:10.1007/s10546-012-9704-y, 2012.

Methane leak detection and sizing over long distances using dual frequency comb laser spectroscopy and a bootstrap inversion technique

Caroline B. Alden[1,2*], Subhomoy Ghosh[3], Sean Coburn[1], Colm Sweeney[2,4], Anna Karion[3], Robert Wright[1], Ian Coddington[3], Kuldeep Prasad[3], Gregory B. Rieker[1]

[1]Department of Mechanical Engineering, University of Colorado at Boulder, Boulder, CO, 80309 USA
[2]Cooperative Institute for Research in Environmental Sciences, Boulder, CO 80309, USA
[3]National Institute of Standards and Technology (NIST), Gaithersburg, MD, 20899, USA
[4]National Oceanic & Atmospheric Administration (NOAA), Boulder, CO 80305, USA

Correspondence to: Caroline B. Alden (caroline.alden@colorado.edu)

Abstract. Advances in natural gas extraction technology have led to increased activity in the production and transport sectors in the United States, and, as a consequence, an increased need for reliable monitoring of methane leaks to the atmosphere. We present a statistical methodology in combination with an observing system for the detection and attribution of fugitive emissions of methane from distributed potential source location landscapes such as natural gas production sites. We measure long (>500 m), integrated open path concentrations of atmospheric methane using a dual frequency comb spectrometer and combine measurements with an atmospheric transport model to infer leak locations and strengths using a novel statistical method, the non-zero minimum bootstrap (NZMB). The new statistical method allows us to determine whether the empirical distribution of possible source strengths for a given location excludes zero. Using this information, we identify leaking source locations (i.e., natural gas wells) through rejection of the null hypothesis that the source is not leaking. The method is tested with a series of synthetic data inversions with varying measurement density and varying levels of model-data mismatch. It is also tested with field observations of 1) a non-leaking source location and 2) a source location where a controlled emission of ~~2~~3.1 E-5 kg s^{-1} of methane gas is released over a period of several hours. This series of synthetic data tests and outdoor field observations using a controlled methane release demonstrate the viability of the approach for the detection and sizing of very small ~~(<2 g m^{-1})~~ leaks of methane across large distances (4+ km^2 in synthetic tests). The field tests demonstrate the ability to attribute small atmospheric enhancements of ~~18~~ 17 ppb to the emitting source location against a background of combined atmospheric (e.g., background methane variability) and measurement uncertainty of 5 ppb (1-sigma), when measurements are averaged over 2 minutes. The results of the synthetic and field data testing show that the new observing system and statistical approach greatly decreases the incidence of false alarms (that is, wrongly identifying a well site to be leaking) compared with the same tests that don't use the NZMB approach, and therefore offers increased leak detection and sizing capabilities.

1 Introduction

The combustion of natural gas in high-efficiency power cycles is cleaner and produces less climate-warming carbon dioxide gas than the combustion of coal (Environmental Protection Agency, 2015), which has led to interest in natural gas as a cleaner alternative to coal for energy generation. Advances in natural gas extraction technology have led to a 35% increase in total natural gas production between 2005 and 2013 in the United States (U.S. Energy Information Administration, 2015). Production is expected to increase by 45% above 2013 levels by the year 2040 (U.S. Energy Information Administration, 2015). A caveat to the promise of natural gas as a lower climate impact energy source, however, is that leaks of methane during extraction and delivery can result in climate warming. Methane gas has high global warming potential (GWP): much higher, for example, than carbon dioxide (CH_4 has a GWP of 28 over 100 years, compared with CO_2, which has GWP of 1 by definition (Myhre et al., 2013)). Above a low threshold (estimated to be ~~≈~~ 3.2% by ~~(~~Alvarez et al.~~,~~ (2012)) leak rate from well to power plant, the near-term climate impacts of using natural gas for power generation become worse than coal (Alvarez et al., 2012; Hayhoe et al., 2002). Recent system-wide analysis suggests that natural gas sector leak rates are likely higher than inventory estimates (Brandt et al., 2014; Zavala-Araiza et al., 2015a). To achieve the lower climate impacts and greater economic benefits of domestic natural gas production, it is important to find low cost methods to detect and reduce methane leakage (Alvarez et al., 2012).

The current industry practice for leak detection and repair (LDAR) is to perform infrequent (annual or less for most sites) "spot" checks for leaks, for example by visual inspection with an optical gas imaging (OGI) camera. However, recent work has shown that methane concentrations measured by OGI cameras can be drastically underestimated if conditions are not ideal, for example under conditions of lower~~if~~ temperature values~~ are cold~~ or~~,~~ higher wind speeds~~ are high~~, or if viewing distances are greater than 50 m (Ravikumar et al., 2016). Furthermore, spot check monitoring is inadequate for detection of leaks, given strong evidence for intermittency of leaks (Allen et al., 2013, 2015a; Mitchell et al., 2015; Subramanian et al., 2015). It has been observed that a small number of facilities leaking at very high rates – so-called "super-emitters" (Frankenberg et al., 2016; Rella et al., 2015; Zavala-Araiza et al., 2015b) – can account for a majority of total emissions (Allen et al., 2013, 2015a, 2015b; Brandt et al., 2014). These characteristics underscore the importance of continuous monitoring for leaks over large areas. Field campaigns with sophisticated atmospheric sampling techniques provide valuable snapshots of the state of natural gas development facility leaks (e.g. Brantley et al., 2014; Karion et al., 2013), but it would be too costly to employ such measurement strategies for long-term continuous monitoring of most natural gas sector facilities.

We present and test an atmospheric measurement system coupled with a statistical inversion approach for detecting and quantifying emissions of methane. The statistical approach is focused on limiting the occurrence of false-positive leak detection. The measurement system used to test the statistical approach is composed of a long-range open-path laser situated in the center of a field of well sites, and a series of retroreflectors around the perimeter of the field to direct light back to a

detector co-located with the laser. The concentration of trace gases along the open beam path (defined as the path between the spectrometer-detector system and a retroreflector) ~~.~~ is determined from the species-specific absorption of light (Dobler et al., 2015; Flesch et al., 2004; Groth et al., 2015; Hashmonay et al., 1999; Levine et al., 2016). Many open-path absorption methods for determining species concentration have been demonstrated (Akagi et al., 2011; Dobler et al., 2015; Flesch et al., 2004; Jones et al., 2011; Nikodem et al., 2015; Wagner and Plusquellic, 2016; Wu et al., 2014). Here we use a dual frequency comb spectrometer (DCS): a unique broadband, high-resolution spectrometer that offers very high stability (low drift) and measurement reproducibility ~~(close agreement of measurements under different conditions, with different instruments or in different locations)~~ of the trace gas measurement, so that concentrations can be compared across ~~long periods of time~~ different conditions and times (Coburn et al., n.d.). It was recently demonstrated that two separate dual frequency comb spectrometers stationed side-by-side and measuring the same 1-km outdoor path showed agreement to 0.35% over a two-week period under ambient variations in temperature, pressure and stability (Waxman et al., 2017). In principle, the range of conditions under which two separate dual frequency comb spectrometers should be comparable is much wider than ambient conditions, because the concentration retrieval is largely dependent on the quality of absorption models (which are well-defined under most conditions experienced at Earth's surface). Previous work also demonstrates that this method of atmospheric trace gas measurement does not require regular or traditional calibration (Coburn et al., n.d.; Rieker et al., 2014; Truong et al., 2016; Waxman et al., 2017). Laboratory and initial field measurements made with the dual frequency comb spectrometer indicate extremely ~~low measurement uncertainty~~ high measurement precision (3 ppb or lower) over long (1 km one-way, or 2 km round-trip) pathlengths (Coburn et al., n.d.; Rieker et al., 2014; Truong et al., 2016; Waxman et al., 2017). The combination of low uncertainty and high stability enable new opportunities for detection and sizing of even very small emissions of methane (Coburn et al., n.d.). Furthermore, the demonstration of sensitive methane measurements over kilometer-scale open paths allows for monitoring methane concentrations over large areas such as natural gas production, processing, and distribution sites. While frequency comb measurements have previously been made in laboratory settings, the recent work of (Coburn et al., n.d.) along with the new work shown here demonstrate the viability of dual frequency comb spectroscopy in real-world conditions.

We use the dual frequency comb measurements in a series of synthetic data and field data tests to demonstrate the utility of the observing system and a novel statistical method for accurately locating one or more point sources of methane within a large area (4+ km^2) using distributed measurements of methane concentrations and an atmospheric transport model. Previous studies have used Gaussian plume models with atmospheric measurements of wind conditions and constituent concentrations to detect sources (e.g., Hirst et al., 2004), and past studies have also shown the utility of open-path lasers for measuring across-plume concentrations for use in the detection of emissions (Flesch et al., 1995; McBain and Desjardins, 2005). Here, we present a novel statistical technique applied to source detection and quantification - with the goal of minimizing false positive source identification. The source-attribution method used here is to apply a non-negative least-squares fitting technique to solve for methane flux at a series of potential source locations (e.g., well heads, pads or other components), given a set of atmospheric

observations and knowledge of atmospheric transport (Leuning et al., 2008). The new statistical approach, called the non-zero minimum bootstrap method (NZMB), uses a bootstrapping of model uncertainties to produce an empirical distribution of source strength for a given well site. Specifically, the empirical distribution is obtained by performing multiple atmospheric inversions (or estimates of surface fluxes using atmospheric data) using a set of resampled atmospheric measurements. The NZMB method establishes a criterion by which well sites or facilities are identified as having non-zero methane emissions based on examination of the minimum value of an ensemble of inversions. That is, a potential leak site is positively identified as a source of methane to the atmosphere if the empirical cumulative distribution of likely source strengths (determined with a series of bootstrap operations) does not include a minimum threshold flux such as zero. Similarly, a facility is identified as not leaking if the empirical cumulative distribution of likely source strengths does include the minimum threshold flux (that is, the minimum value of all bootstrap operations is, for example, zero). By defining a specific null value for each potential leak, this approach reduces the incidence of false positive leak identification (the incorrect attribution of a methane source to a non-leaking facility or well), compared with the same tests that do not use the NZMB method (the "non-bootstrap" approach). For comparison, we run the same series of tests with the non-bootstrap approach, which approximates emissions using a single non-negative least-squares fit.

~~This work presents a set of synthetic experiments and a set of field tests with a controlled leak to prove the utility of this novel observing system for the detection and sizing of methane leaks.~~

Synthetic data tests are performed that assess the effects of increasing measurement density (4, 8, 16, 32, and 64 beams), and the effects of increasing model data mismatch (that is, "observation" noise arising from measurement, transport and other uncertainties). Field tests with atmospheric observation data are performed in a 3 km x 2.5 km field site located in north-central Colorado over the course of one day in January, 2017. The meteorological conditions (e.g., wind speed, wind direction, atmospheric stability) on this day are typical of wintertime and annual mean conditions measured near the field site (for example, compared with conditions at nearby weather station KCOLONGM30). Measurements are made along a series of 3 beams extending from a spectrometer in the middle of the domain. ~~Synthetic data tests assess the effects of increasing measurement density (4, 8, 16, 32, and 64 beams), and the effects of increasing model data mismatch (that is, "observation" noise arising from measurement, transport and other uncertainties).~~

We define leak identification success as maximizing the incidences of leaks found, with a minimal occurrence of false positive source identification, enabling quick response to leaks and avoiding costly mobilization of repair teams due to false positive leak identification. The ability to correctly ascertain the absence of a leak is therefore of equal importance to the ability to find leaks for regulatory compliance applications of this method. With the above tests, we therefore seek to determine 1) whether methane point source emissions can be detected and sized under conditions of observational uncertainty (model-data mismatch) and background variation, 2) whether the absence of a leak can be ascertained in an outdoor field setting, 3) whether

the NZMB method allows for leaks to be positively identified under scenarios of greater simulated model-data mismatch uncertainty~~noise~~, compared with the non-bootstrap method, and 4) whether a higher number of observations increases likelihood that the NZMB and non-bootstrap methods can positively identify leaks. The success of the synthetic and field data tests demonstrates the potential of this observing system for continuous monitoring applications, such as for natural gas facilities, and for providing emission source locations and their approximate strengths. The experiments here also demonstrate the potential for this technology to be used for other source estimation and monitoring applications, for example carbon sequestration.

2 Methods

2.1 Gaussian plume atmospheric transport model

In both the synthetic and real data tests, atmospheric transport is simulated using a Gaussian plume model, using Pasquill-Gifford parameterization of plume dispersion in the lateral and vertical directions (Green et al., 1980; Griffiths, 1994; Hanna et al., 1982). Micrometeorology in the boundary layer is a non-trivial source of uncertainty for characterization of atmospheric flow, and the Gaussian Plume model represents a simplified representation of atmospheric transport and dispersion. It is used to characterize the mean state (or steady-state) of source-receptor relationships with a point source as long as the transport time from source to receptor is comparable to the data averaging time (Gifford, 1976; Hirst et al., 2004)~~, and measurements that integrate concentrations across space, such as those presented here, make the simulation of atmospheric flow with a Gaussian plume model more suitable than for measurements that are discrete in space~~. More sophisticated plume (e.g., AERMOD) or stochastic Lagrangian dispersion models (e.g., WindTrax) and stability parameterizations would be expected to provide more robust representations of the wind shear and inhomogeneities in turbulence in the atmospheric surface layer (Flesch et al., 1995; Perry et al., 1994; Wilson and Sawford, 1996). We select the simplified and low computational cost plume model for assessment of the NZMB method as a baseline test, rather than implementing more advanced representations of transport~~, because of its simplicity and low computational cost~~. Future campaigns aimed at quantification of true emissions will benefit from an assessment of the drawbacks inherent in Gaussian plume model characterization of atmospheric transport, or use of a more sophisticated model, particularly for measurements made at short range.

For the synthetic data tests, the choice of transport model is largely trivial, given that the transport is considered "perfect". Field data is collected with a constant methane source to the atmosphere and a measurement averaging time that is comparable to the source-to-receptor travel time, such that ~~measurement frequency, lateral distance from leak location to beams, and constancy of leak rate make~~ the Gaussian plume model is a simplified but appropriate choice of transport model (Gifford, 1976; Hirst et al., 2004). Because the purpose of this~~e~~ study ~~here~~ is to confirm or reject the basic methodology and not to investigate the impacts of micrometeorological representation on the ability to estimate fluxes, we find the plume model to be sufficient as a baseline test (see Sect. 6).

Neglecting influence of background methane concentrations, Eq. (1) shows the relationship between fluxes and atmospheric concentrations (e.g., Leuning et al., 2008):

$$\mathbf{c} = \mathbf{x} * (\mathbf{c}/\mathbf{x})_{modeled}\ , \qquad (1)$$

Where the $n \times 1$ vector **c** is the atmospheric concentration of the constituent of interest ~~at various points in space~~, and n is the number of measurements. The vector **x** is $m \times 1$ ~~surface~~ sources of the constituent (flux units), where the size of m is equal to the number of potential source flux locations. Here, the vector of fluxes, **x**, is the emission rate of methane from each potential source location. In the synthetic tests and field tests described here, multiple measurements are made on each beam, such that n is always greater than m. The value $(\mathbf{c}/\mathbf{x})_{\text{modeled}}$ is ~~an influence function~~the transport operator matrix describing the relationship between the point source emission~~s~~ and concentrations at observation points (spectrometer beams) under different meteorological conditions, derived using the Gaussian plume model~~. The matrix $(\mathbf{c}/\mathbf{x})_{\text{modeled}}$ is the transport operator matrix representing source-receptor relationships,~~ , and commonly written as **H** (that convention will be followed here)~~ (see Sect. 2.5.3 for details on scaling from point source, to point concentration, to line-averaged concentration).

2.2 Dual Frequency Comb spectrometer for long-range open path methane detection

Dual frequency comb spectrometer measurements are made by transmitting light from the spectrometer through open air at a discrete set of wavelengths where methane absorbs light. The light is transmitted in the direction of a retroreflector, which can be placed 1+ km away (Coburn et al., n.d.; Rieker et al., 2014; Truong et al., 2016; Waxman et al., 2017). The retroreflector directs light back toward a detector co-located with the spectrometer. The amount of light that is absorbed by methane yields a direct measurement of the average ~~mole fraction~~concentration of methane along the open path from spectrometer to retroreflector. The measurements presented here are part of the first campaign to measure atmospheric concentrations with a fielded dual frequency comb spectrometer (Coburn et al., n.d.). The temporal resolution of measurements is related to averaging time: as averaging time increases, measurement precision increases, until such time that atmospheric CH_4 variability begins to erode measurement repeatability (see Sect. 4.1). The spatial resolution of the measurement depends upon beam length, which is easily adjusted by moving retroreflectors closer to or further away from the spectrometer; and beam width, which scales with telescope diameter.

2.3 Flux estimation with Non-Negative Least-Squares fitting solution

We use the Non-Negative Least-Squares (NNLS) algorithm in Fortran-90 to solve for a flux rate (that is, the emission rate from each potential source location), given atmospheric observations (synthetic or real) and atmospheric transport influence functions (Lawson and Hanson, 1995). This algorithm iteratively solves for the best-fit $m \times 1$ vector of fluxes, **x** (see Sect. 2.1 for a description of **x**), given an $n \times 1$ vector of data measurements, **y**, and an $n \times m$ matrix of influence functions, **H**. Given **H**

and **y**, the NNLS algorithm ~~attempts~~ computes a vector **x** (i.e., methane emission rate at each well site) that~~to~~ solves the least squares problem ~~for the vector **x** (i.e., methane emission rate at each well site)~~:

$$\mathbf{Hx} = \mathbf{y}, \text{ subject to } \mathbf{x} >= 0 \quad (2)$$

Uncertainties in **x** and **y** are not included in the NNLS fit; model-data mismatch is used only in generation of the synthetic observations, and not as a control on the solution for **x**. The NNLS algorithm not only returns the solution vector, **x**, but also allows for the calculation of **Hx**, an $n \times 1$ vector describing the expected atmospheric concentration given **H** and the solution for **x**.

2.4 Non-Zero Minimum Bootstrap Analysis

The non-zero minimum bootstrap analysis, or NZMB, is a statistical test of the null hypothesis (*Hypothesis*$_0$) that the source strength at a given well site is equal to 0 kg s^{-1}. It is used here to estimate source strengths in both the synthetic and field data tests. Whereas bootstrapping methods and least-squares methods are not novel techniques, and have previously been applied to problems of source strength estimation, we develop the present methodology with the motivation to seek a solution for fluxes in which the incidence of false-positive source attribution is limited (Efron, 1979; Lawson and Hanson, 1995).

For each of m potential source locations, the null hypothesis (*Hypothesis*$_0$) is that there is no methane emission from that potential source location, and the alternative hypothesis (*Hypothesis*$_1$) is that there is a (non-zero) emission from that potential source location:

$$Hypothesis_0 : x_j = 0 \quad (j = 1,\ldots,m)$$

$$Hypothesis_1 : x_j > 0 \quad (j = 1,\ldots,m)$$

Given that model-data mismatch uncertainty is not zero (i.e., there is uncertainty in the exact relationship between atmospheric observations and surface fluxes due to transport, measurement and other uncertainties), it is not expected that the NNLS fit of **Hx** to **y** is exact, although the problem is overdetermined (that is, $n > m$). We therefore use the mismatch between **Hx** and **y** to create an empirical distribution function describing the confidence interval of the fit to the data, and to accept or reject the null hypothesis claim that we have enough evidence to claim that a particular source is not leaking. That is, the empirical fit to the data is used to quantify uncertainties associated with the model-data mismatch (including, for example, instrument and measurement uncertainties, transport uncertainties and model uncertainties) rather than relying on a "bottom-up" estimation of those sources of uncertainty. We rely on the assumption that model-data mismatch uncertainty has an un-biased Gaussian distribution. Although biases in transport or other sources of uncertainty can exist, we suggest that investigation of that contingency is suited for future studies~~This bootstrapped analysis is described in more detail next~~.

The method for employing the bootstrap analysis is as follows. We first solve for surface-to-atmosphere fluxes of CH_4, **x**, using NNLS, as described in Sect. 2.3. Second, for each observation, y_i ($i = 1,\ldots,n$), we calculate the residual values from the fit to the NNLS solution:

$e\epsilon_{Ri} = y_i - \hat{y}_i$, (3)

where $\hat{y}_i$ (i = 1,…,n) are the individual values in the vector **Hx**. The values of $\hat{y}_i$ (i = 1,…,n) are the "predicted" change in atmospheric methane given the NNLS solution for **x**, or the change in atmospheric methane that is simulated by convolving the source-receptor matrix, **H**, with **x**.

The next step in the NZMB method is: for each observation, y_i (i = 1,…,n), we generate 1000 a new estimates of that observationy_i by using Eq. (3) to sampleleing from the vector of the residuals of the fit to the atmospheric data, **e**, (with replacement, meaning a given value can be sampled more than once), and adding that randomly selected $e_i\epsilon_R$ value to the predicted observation value, $\hat{y}_i y_i$, to create y_{bi} (Efron, 1979). That is, for each observation vector, **y**, we create a new vector, $\mathbf{y}_b$ (b denotes a bootstrapped value):

$y_{bi} = \hat{y}_i + e_{bi}$ (4)

We perform this step 1000 times, This step resulting$s$ in 1000 vectors $\mathbf{y}_b$ (b denotes a bootstrapped value), or 1000 different sets of observations of the form $\{y_{b1},\ldots,y_{bn}\}$, where $y_{bi} = \hat{y}_i y_i + e\epsilon_{bi}$.

For the field data, we apply a moving block bootstrap (Künsch, 1989) because residuals of observations made nearer together in time are more likely to be co-representative, whereas residuals of observations made further apart in time are likely to be less representative due to changes in wind conditions and atmospheric stability. We calculate the autocorrelation in time of the residuals resulting from a single non-negative least-squares fit and use for the moving block window length a value two times the lag time at which the autocorrelation falls below the 95% confidence level. As there is no time dimension in the synthetic data case, we do not apply the moving block bootstrap to those cases.

Next, w We use NNLS to solve for **x** for each of the 1000 resampled sets of observations, yielding 1000 individual solutions for **x**. The final step in the NZMB method is to apply the non-zero-minimum criterion to the 1000 bootstrap solutions for each member of **x**. For each possible source location, we find the minimum value from the 1000-member bootstrap analysis is obtained. The non-zero-minimum criterion states that if the minimum bootstrap value for a given well location is 0 kg s^{-1}, then the source location is classified as having a leak rate of 0 kg s^{-1} (i.e., no leak). This criterion establishes, under the null hypothesis, whether or not 0 (< 0 is not possible since a non-negative least-squares fit is used) is included in the domain of the empirical cumulative distribution function with non-zero mass, described by the 1000 solutions for each well site in **x**. If zero is included in this distribution, then the null hypothesis (**x** = 0) cannot be rejected. Conversely, if 0 is not included in the empirical cumulative distribution function for a given well site (x_j), then the null hypothesis can be rejected and it can be assumed that the well site is leaking. We use a large number of bootstrap members (1000) to ensure that the law of large numbers (LLN) is met. LLN justifies that when the number of bootstrap operations is large, the bootstrapped leak mean approaches the estimated leak from the sample (i.e., the bootstrapped leak mean is a consistent estimator of the estimated leak),

and the distribution of the bootstrapped leak approaches the probability distribution of the source strength. Thus, we can claim that the bootstrapped estimator is a good candidate of the estimated leak from the NNLS, and that the empirical cumulative distribution function is an ~~close enough~~ approximation of the true cumulative distribution function.

After having identified which source locations are non-zero sources to the atmosphere (i.e., leaking), the mean leak strength is estimated as the mean of the 1000 bootstrap solutions for that source location. Uncertainty in the strength of the true leak is calculated as the standard deviation of the 1000 bootstrap solutions at the true leak location.

This method requires little additional computational cost over the non-bootstrap NNLS approach, because additional runs of the transport model are not required, only additional ~~inversions~~ NNLS fits using resampling of the observations. The NZMB approach has the benefit of reducing false positive solutions while also gathering information regarding the parameters of the assumed Gaussian distribution.

2.5 Synthetic Data Tests and Results

2.5.1 "True" leak locations and strengths

To prepare synthetic data testing of the NZMB method, we randomly distribute 20 possible leak source locations within a theoretical 2 km x 2 km domain. This is a reasonable approximation of well density based on high-production regions of the western United States (average well density across the Marcellus and Haynesville shale gas plays are 3+ wells km^{-2}). In the synthetic tests, therefore, $m = 20$. Of the 20 well~~s~~ sites in the domain, we simulate a scenario in which 2 source locations are leaking. The "true" leak rate at well site number 6 is 4.5E-5 kg s^{-1} and the "true" l~~source~~ eak rate at well site number 19 ~~has a rate of~~is 3.0E-5 kg s^{-1}. The remaining 18 well sites are assigned "true" leak ~~strengths~~ rates of 0 kg s^{-1} (Fig. 1). The two non-zero~~"true"~~ leak strengths ~~tested here~~ are very small: roughly half the size of the smallest leaks found by Rella et al. (2015) in a survey of oil and natural gas well pads. The height above ground level of each leak is 0 m.

2.5.2 Idealized meteorological conditions for synthetic data tests

The meteorological data used for synthetic data tests ~~represents~~ includes ~~an idealized scenario in which~~ many wind directions and a variety of wind speeds ~~occur~~ during the sampling of each beam in the domain, representing an ideal scenario for the generation of as many unique, or independent measurements of the leak strength as possible. Leak strengths are ~~simulated to be~~ constant through time, such that the time dimension of the meteorology does not need to be considered. This approach assumes that enough time has passed for all meteorological conditions to have occurred during the sampling of each beam, a condition that ~~that~~ eliminates complications in comparing synthetic cases with different beam orientations. The idealized meteorological field applies 216 unique wind conditions to all beams: three wind speeds (2 m s^{-1}, 3 m s^{-1} and 6 m s^{-1}) from 72 directions (from 5º to 360º, in 5º increments). The conditions represent a situation in which, over a long period of time, many

different wind conditions yield a variety of different measurements downwind of emissions. Given the simple beam configuration presented here, which is agnostic of potential source locations, increasing the number of measurement conditions helps to improve the conditioning of the problem (Crenna et al., 2008; Flesch et al., 2009).

2.5.3 Measurement system configuration and synthetic observations

The "synthetic" atmospheric measurements are simulated based on the dual frequency-comb spectrometer observing system ~~composed of a series of retroreflectors, and a light detector collocated with the spectrometer, as~~ described in Sect. 2.2. The spectrometer is located in the center of the domain, at x = 1000 m and y = 1000 m (Fig. 2).

~~A "beam" is defined as the path between the spectrometer-detector system and a retroreflector.~~ Configurations of 4, 8, 16, 32, and 64 beams per spectrometer-detector system are tested. In all beam configurations, retroreflectors are placed at an equal distance (1000 m) from the spectrometer and at equal distances from neighboring retroreflectors (e.g., Fig. 2). The hub-and-spoke beam configuration is a simple and repeatable pattern for comparison of different numbers of beams. The height of the spectrometer and retroreflectors is 3 m above ground level. Figure 2 shows beams, beam end point locations (retroreflectors) and the spectrometer in a case with 16 beams.

"True" atmospheric methane concentrations are simulated by combining knowledge of atmospheric transport with knowledge of "true" sources and measurement (beam segment) locations with Eq. (1).

In order to generate the synthetic measurement data, each beam path is discretized into 100 segments. For each unique wind condition, "true" source fluxes are multiplied by **H** to calculate atmospheric enhancements at each of the 100 points along the beam path. Enhancements due to leaks are calculated independently for each segment of a beam and subsequently averaged for each beam and for each wind condition. This value mimics the actual data output of the spectrometer, which measures the average concentration along the beam length.

The influence functions describing the relationships between each element of **x** and each segment of each beam path for each wind condition, **H**, are created using the Gaussian plume model described in Sect. 2.1, with neutral stability conditions (Pasquill category D). We create a vector of "true" atmospheric values, **c**, using Eq. (1).

In the synthetic tests, the dimensions of n (e.g., the length of the atmospheric concentration vector, **c**) vary along with the number of beams per spectrometer-detector system and the number of meteorological conditions. In the configuration of 4 beams, for example, n = 216 * 4, because each distinct meteorological condition is applied to each beam. In the 8 beam configuration n = 216 * 8, in the 16 beam configuration n = 216 * 16, and so on.

2.5.4 Perturbation of observations with noise equivalent to model-data mismatch uncertainty~~noise~~

Model-data mismatch is the difference between the true atmospheric CH_4 concentration, **c**, and the simulated or measureable atmospheric CH_4 concentration. This difference is expected to be non-zero due, for example, to measurement uncertainty (sampling and instrumental error), transport uncertainty (imperfect knowledge of air flow between source and observation points), and representation error (for example, the assumption that the measured segment of beam appropriately characterizes the atmospheric concentration at the time and space scales that it represents in the model). We assume here that uncertainty due to the imperfectly known background concentration is also part of model-data mismatch uncertainty. We simulate progressively larger levels of model-data mismatch in order to identify differences in model capabilities to locate and size leaks between the NZMB and non-bootstrap methods.

A range of model-data mismatch values are tested with the expectation that both the NZMB and non-bootstrap models will be more likely to locate and source leaks when lower model-data mismatch is added to the data. To simulate different possible magnitudes of model-data mismatch, the simulated true atmospheric concentrations, **c**, are perturbed with random Gaussian noise with mean 0 ppb and standard deviation equal to the following values: 0.1, 0.2, 0.3, 0.4, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5, 6, 7, 8, 9, and 10 ppb, over a 1 km path. Measurement statistical uncertainty~~noise~~ alone is expected to be on the order of 3 ppb or lower for a 1 km path (Rieker et al., 2014)~~, although recent technological developments may drive measurement uncertainties lower (Coburn et al., n.d.)(Coburn et al., in prep)~~. As the results of field tests will show, the range of model-data mismatch values tested are an appropriate approximation of observed uncertainty (Sect. 4.4)~~noise~~. Model-data mismatch uncertainties~~noise~~ are~~is~~ assumed to be uncorrelated, following convention and understanding of the dual frequency comb measurement scheme. In Eq. (54), $\boldsymbol{\epsilon}$ is a vector of model-data mismatch uncertainty corresponding to the vector, **c**. Both vectors are of length n (i =1,…,n), where n is the number of observations, as described in Sect. 2.5.3. The vector **y** contains the synthetic observations, or the true atmospheric concentrations perturbed with measurement noise.

$$\mathbf{y}_i = \mathbf{c}_i + \boldsymbol{\epsilon}_i \quad (54)$$

2.6 Field Data Observations

2.6.1 Description of field deployed dual comb setup

The first measurements from a field deployed dual frequency comb spectrometer are from the NOAA/ESRL Table Mountain Test Facility, 10 km north of Boulder, Colorado (Fig. 3) (Coburn et al., n.d.). The spectrometer is located near the center of a large (≈3 x 2.5 km) flat-topped mesa that rises several meters above the surrounding terrain (see Fig. 3). The dual frequency comb is housed inside of a trailer, with telescope transceiver affixed to a rotating gimbal on the trailer roof (roughly 4 meters above ground level). The actual dual frequency comb spectrometer is contained in a 56 x 56 x 61-cm electronics rack, and the large trailer provides a field deployment home base. The beam transceiver system sends light between 1620 and 1680 nm, with discrete line spacing of 0.002 nm, through a 2-inch telescope. Dual-comb spectroscopy uses a large spectral bandwidth and

high spectral resolution, which allows for the simultaneous fitting of the absorption pattern for each gas, so that interference among gases is avoided. Background infrared light does not affect the laser signal due to the heterodyne nature of the detection – the detected beat signals between the comb teeth are of high frequency whereas background signals (for example from solar radiation) are of lower frequency. The system emits and senses approximately 28,900 individual comb teeth (Coburn et al., n.d.; Rieker et al., 2014). The wavelength "window" to which the instrument at Table Mountain is tuned is ≈50 nm, spanning 625 individual CH_4 features, 2,482 CO_2 features, and 133 H_2O features. Intensity feedback, triggered data acquisition, and onboard phase correction are quasi-autonomous, enabling the system to operate continuously for any length~~long periods~~ of time (Coburn et al., n.d.; Truong et al., 2016; Waxman et al., 2017).

2.6.2 Leak location and strength

For the field experiments at Table Mountain, a cylinder of compressed methane gas is placed roughly 528 m away from the spectrometer (Fig. 3) with the gas outlet 1 meter above ground level. The methane cylinder is outfitted with a regulator and an Alicat mass flow controller (MC-20SLPM-D). The flow controller is set to release methane in a controlled flow of ~~2~~3.1 E-5 kg s^{-1} at source location 1, between 10:08 and 16:30 on January 26, 2017. The flow rate at source location 2 i~~wa~~s set to 0.0 kg s^{-1} through the duration of January 26, 2017. The controlled methane release point is roughly 0.43 cm in diameter, and the velocity of gas exiting the tubing is negligible.

The field tests are arranged so as to approximate the synthetic tests as closely as possible~~:~~ t~~T~~o emulate the "perfect" background condition of the synthetic tests, the background methane concentration for each source location is measured directly by an upwind beam (Crenna et al., 2008; Flesch et al., 2009). Because the background is assumed to be unique for each source location, each inversion includes only that source location in its solution for fluxes. That is, one inversion is performed for source location 1, and a separate inversion is performed for fluxes at source location 2. The dimensions of m for each test are, therefore, equal to 1, and the dimensions of n for each test are equal to the number of measurements made downwind of the source location.

2.6.3 Retroreflector locations

Three corner-cube retroreflectors ~~("retros")~~ are located near source locations 1 and 2 at Table Mountain (see Fig. 3). At their nearest points, the lateral distances between beams 1 and 2 and source location 1 are 11 m, and 6 m, respectively. The minimum lateral distances between leak location 2 and beams 2 and 3 are 12 m and 8 m, respectively. The horizontal distance from the spectrometer to each retroreflector is 584 m, 585 m, and 588 m, respectively for retroreflectors 1, 2 and 3. All retroreflectors are positioned 1 meter above ground level.

2.6.4 Meteorology at Table Mountain

Wind speed and wind direction are measured directly with a 3D Sonic Anemometer (RM Young 81000 Ultrasonic 3D Anemometer with manufacturer-specified accuracy of ± 0.05 m s^{-1}) located mid-way between the spectrometer and the retroreflectors. It is possible that local wind circulation could lead to meteorological conditions that are not homogenous across the Table Mountain site, which could cause the mean winds measured at the anemometer to not perfectly represent those influencing the plume. Measurement of the entire wind field is not practical, however, so the point measurement is used to characterize meteorology across the site. The suitability of the Gaussian plume model for short-range simulations decreases under low speeds, so all data taken at wind speeds below 0.8 m s^{-1} was removed from this analysis (the reliability of the Gaussian plume model erodes as wind speeds decrease below ≈1 m s^{-1}) (e.g., De Visscher, 2013).

2.6.5 Measurements

We test the bootstrap methodology using measurements taken over the course of one day in January 2017. We test the ability of the bootstrap approach to both disprove the null hypothesis (i.e., to correctly ascertain the presence of a non-zero methane emission) and to prove the null hypothesis (i.e., to correctly ascertain the absence of a leak), by gathering measurements along beam paths that bound: 1) source location 1, where methane is released in a controlled flow rate of ~~2~~3.1 E-5 kg s^{-1}, and 2) source location 2, where no methane is released. Quasi-continuous (626 Hz) data acquisition occurs for 2 minutes on each beam. Time-averaging over 2 minutes is performed to maximize gains in measurement precision as well as to average across shorter time scale eddy mixing events. After a measurement is taken, less than 30 seconds elapse while the gimbal moves to focus the beam on the next retroreflector ("retro") in the measurement sequence. The measurement sequence for the time period of study on January 26, 2017 is: retro 1, retro 2, retro 1, retro 3, retro 1, retro 2, and so on. A fourth retroreflector is included in the measurement sequence (leading to a small time delay between measurements made on retro 3 and retro 2), but data from that beam is not analyzed here for simplicity.

In the field tests, the dimensions of n vary along with the number of measurements taken on the beams used in the fit for the methane emission rate vector, $\mathbf{x}$. For the fit to the methane emission rate at source location 1, all data (that is, all 2-minute measurements) gathered on retroreflectors 1 and 2 are used. For the fit to the emission rate at source location 1, all data gathered on retroreflectors 2 and 3 are used. Upwind measurements are used to constrain background, and downwind measurements are used to determine source strength. The dimensions of n are therefore equal to the number of downwind measurements. For the test at source location 1, n = 63 and for the test at source location 2, n = 30. The value of n is smaller at source location 2 because of the sampling pattern described above.

2.6.6 Background CH_4 Estimation

To most closely approximate the synthetic data testing framework in the field environment, we directly sample background CH_4 concentrations upwind of the leak point. The array of beams shown in Fig. 3 "sandwich" each source location. This configuration means that under most wind conditions (wind directions within ≈40º of orthogonal to the beam array in either direction), one beam is situated upwind and one beam is situated downwind of each source location. With this method, we attempt to remove the time-varying CH_4 concentration to which enhancements from discrete near-field emissions are added. While the Table Mountain site is relatively removed from expected anthropogenic and biogenic methane sources, the presence of nearby small livestock and oil and gas operations means that the background methane concentration does vary according to wind direction and through time. The "beam sandwich" approach, of placing beams on either side of each source location, represents a plausible solution to future regional-scale monitoring of many potential emitters.

3 Results of Synthetic Data Tests

3.1 Synthetic source location with and without the NZMB method

We calculate solutions for **x** using NNLS in a single solution without a bootstrap approach for each set of beam configurations and for each model-data mismatch scenario in the synthetic data case. Figure 4 summarizes the findings of each test by categorizing the results into four outcomes: 2 true leaks found with no false positives, 1 true leak found with no false positives, 0 true leaks found with no false positives, and 1 or more true leaks found with 1 or more false positive. The top half of Fig. 4 (for the non-bootstrap method) shows that, of the 5 different beam configurations tested, all result in false positive source locations under every model-data mismatch scenario when a non-bootstrap approach is taken. That is, even with very low model-data mismatch (0.1 ppb) and many beam measurement locations (64), the non-bootstrap method fails to positively identify true leak sources without also generating false positive results. Non-zero solutions are found for source locations where no "true" leak exists.

The bottom half of Fig. 4 shows the results of the same tests, using instead the NZMB method for locating leaks. The results show that success in leak detection is much higher using NZMB, compared with the non-bootstrap tests. Indeed, none of the NZMB tests result in the occurrence of a false-positive leak location, and only tests with low numbers of beams relative to the number of source locations (4 and 8 beam cases) fail to find both of the true leaks. The 4-beam case results in positive identification of both leaks up to a model-data mismatch threshold of 2 ppb, above which 1 true leak is found. One leak is consistently found up to a threshold of 5 ppb, and above 5 ppb model data mismatch no true leaks are identified (but no false positives are generated either). The 8-beam case results in accurate location of both true leaks up to a model-data mismatch threshold of 3.5 ppb, above which 1 true leak is found (with no false positives). One leak is consistently found up to the maximum testing point of 10 ppb. In order to consistently locate both true leaks with no false positive results under all model-

data mismatch scenarios, 16 or more beams are needed for the set of cases that are tested here. Alternate configurations of "true" leaks at well sites other than 6 and 9 are not tested, however given that meteorological conditions are simulated equally from all directions, we would not expect a different set of results from a different set of "true" leaks.

A subset of the~~The~~ results for the 8-beam NNLS without bootstrap and the NNLS with NZMB cases are shown in Figs. 5 and 6 (for conciseness; all results are shown in the Supporting Information). It is evident from Fig. 5 that, even with very low model-data mismatch noise (0.~~1~~5 ppb), the non-bootstrap model results in well sites other than the 2 true leak locations being erroneously identified as sources of methane. It is evident from Fig. 5 that, as model-data mismatch increases, the strength of incident false positive results also increases. By contrast, no false-positive leaks are identified in the NZMB case shown in Fig. 6, at any level of model-data mismatch noise. Above a model-data mismatch threshold of 4 ppb, only one of two true leaks ~~are~~ is found in the 8-beam case using NSMB. As Fig. 4 shows, 16 or more beams are necessary to consistently find both true leaks at higher thresholds of model-data mismatch uncertainty using the NZMB method, given the hub-and-spoke beam placement scheme tested here. More complex placement of beams (for example placing beams closer to known well sites) would likely result in even better ability to locate leaks with fewer beams.

3.2 Synthetic source sizing using the NZMB method

Synthetic data tests of the new bootstrap methodology presented here show high success in leak location, with zero incidence of false positive leak detections. ~~Following successful identification of leak locations with the NZMB method, leak sizes can be estimated using the statistical distribution of bootstrap operations.~~ Figure 6 shows the maximum and minimum values of 1000 bootstrap operations for each model-data mismatch test case for the 8-beam configuration. At low levels of model-data mismatch uncertainty (0.1-0.5 ppb), the maximum and minimum solutions bound a small range that is close to the true leak strength. As higher levels of model-data mismatch noise are added to observations, the maximum and minimum values diverge. However, even as the maximum and minimum solutions diverge, most cases include the true leak strength within the maximum and minimum bounds. ~~As described in Sect. 2, we use the mean and standard deviation of 1000 operations to estimate leak flow rate and uncertainty.~~

Using the NZMB method, all beam cases (even the 4-beam case) correctly identify that both well sites 6 and 19 are emitting methane when model-data mismatch is 2 ppb or lower (Fig. 4). At that level of model-data mismatch, higher numbers of beams and observations tend to lead to lower standard deviation around the mean estimated leak strength and a more accurate estimate of true leak strength (Table 1). An exception is at well site 19, where the 8-beam case did not perform as well as the 4-beam case. It may be that both cases were inadequate for accurately sizing leaks, and that 16 beams are necessary in a dense field of wells such as is tested here. The failure of the 8-beam case to accurately predict the leak rate at well site 19 is also evident from histograms of bootstrap operations, shown for each beam case with model-data mismatch of 2 ppb in Fig. 7.

Histograms of the results for the 16, 32 and 64 beam cases with 10 ppb model-data mismatch are shown in Fig. 8. It is clear from Fig. 8 that, even with very high model-data mismatch uncertainty, simple hub-and-spoke configurations of between 16-64 beams are able to locate and estimate leak flow rates to within reasonable bounds of uncertainty.

4 Results of Field Data Tests

4.1 Performance overview of field deployed DCS

Atmospheric observations are made over the course of one day on January 26, 2017 at the Table Mountain site. A set of 3 retroreflectors create long-range open-path beams of ≈585 m (Fig 3). Spectrometer performance in the field demonstrates no loss of precision or reliability compared with laboratory performance, as demonstrated by (Coburn et al., n.d.). Figure 9 shows a plot of Allan deviations for January 26, 2017, demonstrating measurement precision of 5-6 ppb when measurements are averaged for 2 minutes. Precision of field measurements is limited by repeatability of measurements and atmospheric variability of CH_4 because measurements are time-averaged; the latter is likely a dominant driver of uncertainty in this case, as will be discussed in Sect. 4.3. The Allan deviation in Fig. 9 shows improvement of precision with averaging time, to a minimum at ≈70 seconds, followed by an increase that is likely due to atmospheric variability.

4.2 Atmospheric Observations of CH_4 at Table Mountain

On January 26, 2017, measurements are made throughout the day, including during a 6.5 hour controlled release of methane at ~~one of two local~~ source location~~s~~ ~~(~~1~~)~~. At ~~an~~ adjacent source location ~~(~~2~~)~~, no methane release is emitted. A series of 3 retroreflectors are oriented such that each source region is monitored independently from the other; one beam on either side of each source location serves as a "background" measurement. We examine the results of two separate inversion tests: 1) a day-long set of observations of ~~the~~ source location 1 (with the controlled release) that is situated between retroreflectors 1 and 2, and 2) a day-long set of observations of ~~a~~ non-leaking source location 2 that is situated between retroreflectors 2 and 3. These tests were performed simultaneously, such that contamination from source location 1 could have resulted in background contamination for monitoring of source location 2.

On January 26, 2017, mean wind speeds are 2.1 m/s and winds are primarily from the east and northeast, so that retroreflector 1 is downwind of the controlled release, and retroreflector 2 is upwind of the controlled release. Similarly, retroreflector 2 is downwind of ~~the~~ non-leaking source location 2 and retroreflector 3 is upwind of ~~the~~ non-leaking source 2 (Fig. 3). Stability Classes ranged from B (moderately unstable) to D (neutral) throughout the course of the day (see Supplemental Information for a timeseries of stability and detailed description of its calculation). We use the ~~(~~Griffiths~~,~~ (1994) corrections to the ~~(~~Briggs~~,~~ (1974) parameterizations to calculate σ_y and σ_z.

At source location 1 (Fig. 10, panel a), during the period when the controlled release is on (non-zero flow), the downwind retroreflector (Retro 1) shows a clear enhancement above the concentration measured on the upwind retroreflector (Retro 2), except during the middle of the day when the winds shift briefly to the south (Fig. 10, panel c). The mean of all CH_4 measurements along beam 1 during the period that the leak is on is 2046~~3~~ ppb; the mean CH_4 measured along beam 2 during the same period is 2025~~6~~ ppb. Both retroreflectors demonstrate changes in background CH_4 concentrations over the course of the day; the range in values measured on the upwind retroreflector is 658~~2~~ ppb. There ~~appears to~~may be a~~some~~ relationship between ambient CH_4 concentration and wind direction, as both retroreflectors show a drastic decrease in concentration when the winds abruptly shift to the West at 16:30 (which happens to coincide with the time the leak was turned off).

At source location 2, no leak is released during the period of study, and throughout the course of the day, both retroreflectors 2 and 3 measure similar changes in atmospheric CH_4 variability (Fig. 10 panel b). The range of measured values over the course of the entire day are 128 ppb on beam 2 and 124~~6~~ ppb on beam 3. The mean of all CH_4 measurements (throughout the course of the day) is 2016~~5~~ ppb on beam 2, and 2019~~20~~ ppb on beam 3.

4.3 Background CH_4 Observations

The beams stationed upwind of each source location provide estimates of the background CH_4 concentration inflow for that site. ~~At source location 1, the winds are such on January 26th that retro 2 is upwind, providing background concentration measurements for that location. At source location 2, retro 3 is the upwind or background concentration beam. We apply a tolerance criteria of 5 minutes, so that for a measurement on beam 2 (3) to be used as the background for beam 1 (2), no more than 5 minutes can have elapsed between the middle points of the two measurement times.~~ After ~~each set of~~a linear interpolation to upwind ~~and downwind~~ measurements has been applied~~been paired within this time window~~, this is background~~e concentration measured on the upwind beam~~ is subtracted from ~~that~~ measurements~~d~~ on the downwind beam, to yield a measure of the CH_4 enhancement due to fluxes at the source location. Applying this method, the mean and standard deviation of the enhancement above background on retroreflector 1 – which is downwind of source location 1 (leak rate of ~~2~~3.1E-5 kg s^{-1}) – is ~~18~~17.~~0~~4 ± 10.1 ppb. Applying this method to source location 2, we find the~~a~~ mean and standard deviation of the enhancement on retroreflector 2 – which is downwind of source location 2 (leak rate of 0 kg s^{-1}) – is~~of 0~~3.~~5~~1 ± 7.3 ppb, a value ~~well~~ within the range of variability expected from combined measurement and background uncertainty ~~and background variability~~.

4.4 Field-based estimates of model-data mismatch

We examine measurements at source location 2 in order to estimate model-data mismatch in the field, for comparison with the model-data mismatch values applied in the synthetic data tests. By examining the difference between measurements made on different retroreflectors (retroreflectors 2 and 3) at similar points in time (within 5 minutes), we obtain an approximation of the contributions to model-data mismatch arising from measurement uncertainty, ~~-~~representation uncertainties, ~~and~~ background ~~variability~~ construction (that is, the method of background estimation) and background sampling (that is, the method of

sampling background concentrations) on the timescale relevant for estimation using this technique. We find a standard deviation of ~~6~~ 5 ppb~~, which suggests that the range of model-data mismatch values tested in the synthetic data experiments are appropriate~~. This value differs from the standard deviation of the enhancement for the entire timeseries (reported above in Sect. 4.3) because it compares differences in upwind and downwind concentrations measured at approximately the same time. We add (in quadrature) an estimate of the transport uncertainty that includes uncertainties in measurement of wind speed and wind direction, atmospheric stability parameterization, and placement of the sonic anemometer relative to the leak location (see Supplemental Information for detail). The method for estimating transport uncertainty yields transport uncertainty arising from a plume that interacts with any location along the beam, and therefore requires knowledge of the mean distance between the leak point and each segment of the beam. The estimated transport uncertainty, calculated in this way, is 0.8 ppb. If, for example, the wind direction is perfectly perpendicular to the beam for the entirety of the measurement period (which does not occur on January 26th, 2017), then the leak-to-beam distance used in the calculation should collapse to the minimum lateral distance between the leak and the beam. Using that value instead, transport uncertainty is 12.2 ppb. The overall value of model-data mismatch (reflecting combined measurement, background, and transport uncertainty), estimated in this way, is therefore 5.1 ppb with a maximum range of 13.2 ppb, which suggests that the range of model-data mismatch values tested in the synthetic data experiments are appropriate. The Allan deviation in Fig. 9 shows a similar level of measurement uncertainty, which suggests that most of the uncertainty observed in our record is captured in this estimate of model-data mismatch, which includes effects of atmospheric variability. Precision could be improved by averaging data over a shorter time span (70 seconds), but those gains would be minimal (Fig. 9).

4.5 Results of inversions using Table Mountain observations

Both the Non-bootstrap and the NZMB approaches accurately predict the presence of methane emissions at source location 1 (Table 2). The average bootstrapped flux value is within ~~1~~2-sigma of the true flux value measured at the flow meter at source location 1 (Fig. 11). At source location 2, the non-bootstrap approach falsely predicts a positive emission rate of 0.5~~2~~ E-5 kg s^{-1} (Table 2) where no leak is present. The NZMB approach, by contrast, is able to accurately predict that there is no leak present at source location 2, because the minimum of the 1000 bootstrap solutions is zero (Fig. 11). As the synthetic data tests demonstrated, the NZMB method is necessary to avoid false identification of leaking source locations. The field data tests corroborate that the new bootstrap approach enables higher confidence of accurate attribution of emissions to source locations without generating “false alarms”.

5 Discussion

The results of this study demonstrate success of the new observing system in finding one or more leaks of methane in a field of wells, using synthetic and field data for confirmation. The methods presented here for locating and sizing leaks of methane

in a field of natural gas production facilities succeeds not only in identifying the location of a leak, but it also does so with no incidences of "false positive" leak detection in either the synthetic or field data tests.

5.1 Synthetic Data Tests

The results of the synthetic data tests demonstrate how the observing system tested in the field for a single source location can be expanded for simultaneous monitoring of many source locations. We find that synthetic tests performed without the NZMB methodology failed to identify the presence of leaks as reliably as synthetic tests performed with the NZMB method, demonstrating the improved robustness of our new statistical methods for leak detection. In the non-bootstrap tests, all synthetic data cases resulted in false positive solutions (Fig. 4). By contrast, the NZMB method succeeds in correctly identifying two leaks of strength 3.0E-5 and 4.5E-5 kg s^{-1} with 4 or more beams monitoring 20 wells in a 4 km^2 area, with 2 ppb model-data mismatch uncertainty (a condition that could conceivably also be met in the field given low background uncertainty and high measurement precision). The NZMB method also consistently succeeds in finding both leaks with 16 or more beams with at least 10 ppb model-data mismatch uncertainty. Notably, the NZMB method locates and sizes both leaks with no false positive results. Determination of leak strength was successful to within 25% (and all but a few cases well below 10%) for all cases with 16 or more beams, using the NZMB method.

5.2 Field Data Tests

Field data testing of the NZMB method corroborates the synthetic data findings: that the new atmospheric observing system presented here results in high accuracy of leak detection without false positive results. The ability of the dual frequency comb spectrometer to identify a very small leak (~~2~~3.1 E-5 kg s^{-1}), relying on very small methane enhancements (~~18~~ 17 ppb) against a highly variable background (range of 65~~82~~ ppb), demonstrates the potential power of this method for the methane leak detection over large areas.

An important caveat to the methodology presented here is the short length of the measurement averaging time, which presents a mismatch with the ideal application of most dispersion models (for which practice is generally to use averaging times longer than 2 minutes). This requirement in our methodology is due to two factors: the first is that rapid scanning for potential leaks is an important feature in areas where many sites must be monitored and leaks can be intermittent. The second factor is that background methane concentrations can vary with high frequency (order minutes) in proximity to areas of oil and natural gas production (Dlugokencky et al., 1995). We attempt to mitigate uncertainties arising from using dispersion parameters developed for longer time-scale modeling over a 2-minute period in several ways. First, *n* 2-minute dispersion calculations gathered over longer time scales (*n* is between 30 and 63 for field data tests shown here) are aggregated for use in a single inversion, which is accepted practice (Scire et al., 2000). Second, both sources and receptors are close to the surface, which may help to mitigate crosswind-integrated concentration fluctuation intensity (Weil et al., 2012). Third, a sensitivity test in which we adjust the horizontal dispersion coefficient, σ_y, for shorter time averaging, using the methods of (Gifford, 1976),

shows negligible changes in the results (Supplemental Information). We find that the potential value of a method for rapid detection of methane emissions over large scales and against a highly variable background means that the uncertainties introduced from modeled eddy diffusivity parameterization are a complicating but not irreconcilable caveat.

6 Conclusions

The focus of this study is to show the powerful potential of the combination of a new statistical method with dual frequency comb spectroscopy for the location and sizing of point source emissions. The synthetic and field tests presented here rely on near-perfect (in the synthetic data tests) or well-constrained (in the field data tests) background concentration estimation. Future studies are needed to address the potential complications of more complex background conditions and meteorological conditions under which it is not possible to obtain sequential "upwind" and "downwind" samples. Similarly, the tests here rely on the assumption of constant leak rates, which is likely not a realistic assumption that can be made for methane emissions from oil and gas operations. Future work to address these complexities will be necessary. Future studies are also needed to examine the gains that can be made from optimization of beam configurations for improved leak detection given variable wind and background conditions. In particular, previous work has shown the critical impact that sensor placement can pose on the conditioning of the source-receptor relationship matrix (**H**), and suggests paths forward for optimization of sensor placement (Crenna et al., 2008; Flesch et al., 2009). Specifically, the placement of one beam or sensor between each source to be apportioned would be expected to lead to a lower condition number and therefore a more reliable result (Crenna et al., 2008; Flesch et al., 2009). Work aimed at addressing these complications is underway, as are inversion efforts to resolve issues of leak intermittency.

A notable aspect of the micrometeorological modeling used here to demonstrate the NZMB methodology is the simple representation of atmospheric transport (the Gaussian plume model). The choice to use a simple model that is familiar to the broader scientific community is intentional, however its use belies the complex nature of turbulent mixing and dispersion in the atmospheric surface layer. What is gained in simplicity and in providing a baseline for the most basic performance of the methodology in a field setting may come at the cost of recommending a model that may not ultimately be well-suited for such an endeavor. The Gaussian plume model neglects important aspects of atmospheric mixing such as wind shear and the height dependence of eddy diffusivity, and better models exist for simulation of atmospheric flow at this scale. It is assumed that more sophisticated models of atmospheric dispersion could, therefore, lead to better flux estimation. We suggest that future applications in field settings of the methodology presented here consider their use. Importantly, despite its drawbacks, the Gaussian plume model proves sufficient in the tests here for the accurate identification (and, importantly, avoidance of misidentification) of controlled, field-based methane leaks. Future studies of the best transport model for the application of DCS measurements and the NZMB method for leak detection is warranted.

The initial work presented here demonstrates the promising potential of dual frequency comb spectroscopy for detection of leaks in the natural gas supply chain, and the valuable gains that can be provided by using the NZMB method over the NNLS fitting technique alone.

The authors declare that they have no conflict of interest.

Author Contribution

S. Ghosh, K. Prasad, and C. Alden implemented the statistical NZMB technique. C. Alden, S. Coburn, R. Wright, C. Sweeney, K. Prasad, S. Ghosh and G. Rieker designed the experiments and S. Coburn, R. Wright, and C. Alden carried them out. A. Karion provided expert guidance and experimental design input. C. Alden prepared the manuscript with contributions from all co-authors.

Data Availability

All data will be made publically available on an FTP server upon publication. This section will be updated with an address to that site.

Acknowledgements

The information, data, or work presented herein was funded in part by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000539. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

Allen, D. T., Torres, V. M., Thomas, J., Sullivan, D. W., Harrison, M., Hendler, A., Herndon, S. C., Kolb, C. E., Fraser, M. P., Hill, A. D., Lamb, B. K., Miskimins, J., Sawyer, R. F. and Seinfeld, J. H.: Measurements of methane emissions at natural gas production sites in the United States, Proc. Natl. Acad. Sci., 110, 17768–17773, doi:10.1073/pnas.1304880110, 2013.

Allen, D. T., Pacsi, A. P., Sullivan, D. W., Zavala-Araiza, D., Harrison, M., Keen, K., Fraser, M. P., Hill, A. D., Sawyer, R. F. and Seinfeld, J. H.: Methane emissions from process equipment at natural gas production sites in the United States: pneumatic controllers, Environ. Sci. Technol., 49(1), 630–640, doi:10.1021/es5040156, 2014.

Allen, D. T., Sullivan, D. W., Zavala-Araiza, D., Pacsi, A. P., Harrison, M., Keen, K., Fraser, M. P., Hill, A. D., Lamb, B. K., Sawyer, R. F. and Seinfeld, J. H.: Methane Emissions from Process Equipment at Natural Gas Production Sites in the United States: Liquid Unloadings, Environ. Sci. Technol., 49, 641–648, doi:dx.doi.org/10.1021/es504016r, 2015.

Alvarez, R. A., Pacala, S. W., Winebrake, J. J., Chameides, W. L. and Hamburg, S. P.: Greater focus needed on methane leakage from natural gas infrastructure, Proc. Natl. Acad. Sci., 109(17), 6435–6440, doi:10.1073/pnas.1202407109, 2012.

Brandt, A. R., Heath, G. A., Kort, E. A., O'Sullivan, F., Petron, G., Jordaan, S. M., Tans, P., Wilcox, J., Gopstein, A. M., Arent, D., Wofsy, S., Brown, N. J., Bradley, R., Stucky, G. D., Eardley, D. and Harriss, R.: Methane Leaks from North American Natural Gas Systems, Science (80-.)., 343, 733–735, doi:10.1126/science.1247045, 2014.

Brantley, H. L., Thoma, E. D., Squier, W. C., Guven, B. B. and Lyon, D.: Assessment of Methane Emissions from Oil and Gas Production Pads using Mobile Measurements, Environ. Sci. Technol., 48, 14508–14515, doi:dx.doi.org/10.1021/es503070q, 2014.

Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., Crounse, J. D. and Wennberg, P. O.: Emission factors for open and domestic biomass burning for use in atmospheric models, Atmos. Chem. Phys., 11(9), 4039–4072, doi:10.5194/acp-11-4039-2011, 2011.

Allen, D. T., Torres, V. M., Thomas, J., Sullivan, D. W., Harrison, M., Hendler, A., Herndon, S. C., Kolb, C. E., Fraser, M. P., Hill, A. D., Lamb, B. K., Miskimins, J., Sawyer, R. F. and Seinfeld, J. H.: Measurements of methane emissions at natural gas production sites in the United States, Proc. Natl. Acad. Sci., 110, 17768–17773, doi:10.1073/pnas.1304880110, 2013.

Allen, D. T., Sullivan, D. W., Zavala-Araiza, D., Pacsi, A. P., Harrison, M., Keen, K., Fraser, M. P., Hill, A. D., Lamb, B. K., Sawyer, R. F. and Seinfeld, J. H.: Methane Emissions from Process Equipment at Natural Gas Production Sites in the United States: Liquid Unloadings, Environ. Sci. Technol., 49, 641–648, doi:dx.doi.org/10.1021/es504016r, 2015a.

Allen, D. T., Pacsi, A. P., Sullivan, D. W., Zavala-Araiza, D., Harrison, M., Keen, K., Fraser, M. P., Hill, A. D., Sawyer, R. F. and Seinfeld, J. H.: Methane emissions from process equipment at natural gas production sites in the United States: pneumatic controllers, Environ. Sci. Technol., 49(1), 633–640, doi:10.1021/es5040156, 2015b.

Alvarez, R. A., Pacala, S. W., Winebrake, J. J., Chameides, W. L. and Hamburg, S. P.: Greater focus needed on methane leakage from natural gas infrastructure, Proc. Natl. Acad. Sci., 109(17), 6435–6440, doi:10.1073/pnas.1202407109, 2012.
Brandt, A. R., Heath, G. A., Kort, E. A., O'Sullivan, F., Petron, G., Jordaan, S. M., Tans, P., Wilcox, J., Gopstein, A. M., Arent, D., Wofsy, S., Brown, N. J., Bradley, R., Stucky, G. D., Eardley, D. and Harriss, R.: Methane Leaks from North American Natural Gas Systems, Science (80-.)., 343, 733–735, doi:10.1126/science.1247045, 2014.

Brantley, H. L., Thoma, E. D., Squier, W. C., Guven, B. B. and Lyon, D.: Assessment of Methane Emissions from Oil and Gas Production Pads using Mobile Measurements, Environ. Sci. Technol., 48, 14508–14515, doi:dx.doi.org/10.1021/es503070q, 2014.
Briggs, G. A.: Diffusion estimation for small emissions, in ATDL Contribution File No. 79, Air Resources Atmospheric Turbulence and Diffusion Laboratory, NOAA, Oak Ridge, Tennessee., 1974.

Coburn, S., Alden, C. B., Wright, R., Cossel, K., Baumann, E., Truong, G.-W., Giorgetta, F., Sweeney, C., Newbury, N. R., Prasad, K., Coddington, I. and Rieker, G. B.: Continuous regional trace gas source attribution using a field-deployed dual frequency comb spectrometer, submitted, n.d.

Crenna, B. P., Flesch, T. K. and Wilson, J. D.: Influence of source – sensor geometry on multi-source emission rate estimates, Atmos. Environ., 42, 7373–7383, doi:10.1016/j.atmosenv.2008.06.019, 2008.

Dlugokencky, E. J., Steele, L. P., Lang, P. M. and Masarie, K. A.: Atmospheric methane at Mauna Loa and Barrow

observatories ' Presentation and analysis of in situ measurements Edward, J. Geophys. Res., 100(Dll), 23103–23113, 1995.

Dobler, J., Zaccheo, T. S., Blume, N., Braun, M., Botos, C. and Pernini, T. G.: Spatial mapping of greenhouse gases using laser absorption spectrometers at local scales of interest, Proc. SPIE, 9645, 96450K1-9645K13, doi:10.1117/12.2197713, 2015.

Efron, B.: Bootstrap methods: Another look at the jackknife, Ann. Stat., 7(1), 1–26, doi:10.1214/aos/1176344552, 1979.

Environmental Protection Agency: Inventory of U.S. greenhouse gas emissions and sinks: 1990-2013, Fed. Regist., 80(36), 9718, doi:EPA 430-R-13-001, 2015.

Flesch, T. K., Wilson, J. D. and Yee, E.: Backward-Time Lagrangian Stochastic Dispersion Models and Their Application to Estimate Gaseous Emissions, J. Appl. Meteorol., 34(6), 1320–1332, doi:10.1175/1520-0450(1995)034<1320:BTLSDM>2.0.CO;2, 1995.

Flesch, T. K., Wilson, J. D., Harper, L. A., Crenna, B. P. and Sharpe, R. R.: Deducing Ground-to-Air Emissions from Observed Trace Gas Concentrations: A field trial, J. Appl. Meteorol., 43, 487–502, doi:10.1175/1520-0450(2004)043<0487:DGEFOT>2.0.CO;2, 2004.

Flesch, T. K., Harper, L. A., Desjardins, R. L., Gao, Z. and Crenna, B. P.: Multi-Source Emission Determination Using an Inverse-Dispersion Technique, Boundary-Layer Meteorol., 132, 11–30, doi:10.1007/s10546-009-9387-1, 2009.

Frankenberg, C., Thorpe, A. K., Thompson, D. R., Hulley, G., Kort, E. A., Vance, N., Borchardt, J., Krings, T., Gerilowski, K., Sweeney, C., Conley, S., Bue, B. D., Aubrey, A. D., Hook, S. and Green, R. O.: Airborne methane remote measurements reveal heavy-tail flux distribution in Four Corners region, Proc. Natl. Acad. Sci., 113(35), 9734–9739, doi:10.1073/pnas.1605617113, 2016.

Gifford, F. A.: Atmospheric dispersion models for environmental pollution, in Lectures on Air Pollution and Environmental Impact Analysis, edited by D. A. Haigen, pp. 35–58, Boston, Mass., 1976.

Green, A. E. S., Singhal, R. P. and Venkateswar, R.: Analytic Extensions of the Gaussian Plume Model, J. Air Pollut. Control Assoc., 30(7), 773–776, doi:10.1080/00022470.1980.10465108, 1980.

Griffiths, R. F.: Errors in the use of the Briggs parameterization for atmospheric dispersion coefficients, Atmos. Environ., 28(17), 2861–2865, doi:doi:10.1016/1352-2310(94)90086-8, 1994.

Groth, A., Maurer, C., Reiser, M. and Kranert, M.: Determination of methane emission rates on a biogas plant using data from laser absorption spectrometry, Bioresour. Technol., 178, 359–361, doi:10.1016/j.biortech.2014.09.112, 2015.

Hanna, S. R., Briggs, G. A. and Hosker, Rayford P, J.: Handbook on atmospheric diffusion., 1982.

Hashmonay, R. a., Yost, M. G., Mamane, Y. and Benayahu, Y.: Emission rate apportionment from fugitive sources using open-path FTIR and mathematical inversion, Atmos. Environ., 33(5), 735–743, doi:10.1016/S1352-2310(98)00228-3, 1999.

Hayhoe, K., Kheshgi, H. S., Jain, A. K. and Wuebbles, D. J.: Substitution of natural gas for coal: Climatic effects of utility sector emissions, Clim. Change, 54, 107–139, doi:doi:10.1023/A:1015737505552, 2002.

Hirst, B., Gibson, G., Gillespie, S., Archibald, I., Podlaha, O., Skeldon, K. D., Courtial, J., Monk, S. and Padgett, M.: Oil and gas prospecting by ultra-sensitive optical gas detection with inverse gas dispersion modelling, Geophys. Res. Lett., 31, 1–4,

doi:10.1029/2004GL019678, 2004.

Jones, F. M., Phillips, F. A., Naylor, T. and Mercer, N. B.: Methane emissions from grazing Angus beef cows selected for divergent residual feed intake, Anim. Feed Sci. Technol., 166–167, 302–307, doi:10.1016/j.anifeedsci.2011.04.020, 2011.

Karion, A., Sweeney, C., Pétron, G., Frost, G., Michael Hardesty, R., Kofler, J., Miller, B. R., Newberger, T., Wolter, S., Banta, R., Brewer, A., Dlugokencky, E., Lang, P., Montzka, S. a., Schnell, R., Tans, P., Trainer, M., Zamora, R. and Conley, S.: Methane emissions estimate from airborne measurements over a western United States natural gas field, Geophys. Res. Lett., 40(16), 4393–4397, doi:10.1002/grl.50811, 2013.

Künsch, H. R.: The jackknife and the bootstrap for general stationary observations, Ann. Stat., 17(3), 1217–1241 [online] Available from: http://www.jstor.org/stable/2241719, 1989.

Lawson, C. L. and Hanson, R. J.: Solving Least Squares Problems, Prentice-Hall, Jet Propulsion Laboratory., 1995.

Leuning, R., Etheridge, D., Luhar, A. and Dunse, B.: Atmospheric monitoring and verification technologies for CO2 geosequestration, Int. J. Greenh. Gas Control, 2(3), 401–414, doi:10.1016/j.ijggc.2008.01.002, 2008.

Levine, Z. H., Pintar, A. L., Dobler, J. T., Blume, N., Braun, M., Zaccheo, T. S. and Pernini, T. G.: The detection of carbon dioxide leaks using quasi-tomographic laser absorption spectroscopy measurements in variable wind, Atmos. Meas. Tech., 9(4), 1627–1636, doi:10.5194/amt-9-1627-2016, 2016.

McBain, M. C. and Desjardins, R. L.: The evaluation of a backward Lagrangian stochastic (bLS) model to estimate greenhouse gas emissions from agricultural sources using a synthetic tracer source, Agric. For. Meteorol., 135(1–4), 61–72, doi:10.1016/j.agrformet.2005.10.003, 2005.

Mitchell, A. L., Tkacik, D. S., Roscioli, J. R., Herndon, S. C., Yacovitch, T. I., Martinez, D. M., Vaughn, T. L., Williams, L., Sullivan, M., Floerchinger, C., Omara, M., Subramanian, R., Zimmerle, D., Marchese, A. J. and Robinson, A. L.: Measurements of Methane Emissions from Natural Gas Gathering Facilities and Processing Plants: Measurement Results, Environ. Sci. Technol., 49(20), 12602, doi:10.1021/acs.est.5b04018, 2015.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T. and Zhang, H.: Anthropogenic and Natural Radiative Forcing, in Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2013.

Nikodem, M., Plant, G., Sonnenfroh, D. and Wysocki, G.: Open-path sensor for atmospheric methane based on chirped laser dispersion spectroscopy, Appl. Phys. B Lasers Opt., 119(1), 3–9, doi:10.1007/s00340-014-5938-3, 2015.

Perry, S., Cimorelli, A., Lee, R., Paine, R., Venkatram, A., Weil, J. and Wilson, R.: AERMOD: a dispersion model for industrial source applications, Washington, D.C., 1994.

Ravikumar, A. P., Wang, J. and Brandt, A. R.: Are Optical Gas Imaging Technologies Effective For Methane Leak Detection?, Environ. Sci. Technol., acs.est.6b03906, doi:10.1021/acs.est.6b03906, 2016.

Rella, C. W., Tsai, T. R., Botkin, C. G., Crosson, E. R. and Steele, D.: Measuring emissions from oil and natural gas well pads using the mobile flux plane technique, Environ. Sci. Technol., 49(7), 4742–4748, doi:10.1021/acs.est.5b00099, 2015.

Rieker, G. B., Giorgetta, F. R., Swann, W. C., Kofler, J., Zolot, A. M., Sinclair, L. C., Baumann, E., Cromer, C., Petron, G., Sweeney, C., Tans, P. P., Coddington, I. and Newbury, N. R.: Frequency-comb-based remote sensing of greenhouse gases over kilometer air paths, Optica, 1(5), 290–298, doi:10.1364/OPTICA.1.000290, 2014.

Scire, J. S., Strimaitis, D. G. and Yamartino, R. J.: A User's Guide for the CALPUFF Dispersion Model, Concord, MA., 2000.

Subramanian, R., Williams, L. L., Vaughn, T. L., Zimmerle, D., Roscioli, J. R., Herndon, S. C., Yacovitch, T. I., Floerchinger, C., Tkacik, D. S., Mitchell, A. L., Sullivan, M. R., Dallmann, T. R. and Robinson, A. L.: Methane emissions from natural gas compressor stations in the transmission and storage sector: Measurements and comparisons with the EPA greenhouse gas reporting program protocol, Environ. Sci. Technol., 49(5), 3252–3261, doi:10.1021/es5060258, 2015.

Truong, G.-W., Waxman, E. M., Cossel, K. C. C., Baumann, E., Klose, A., Giorgetta, F. R., Swann, W. C., Newbury, N. R. and Coddington, I. C.: Accurate frequency referencing for fieldable dual-comb spectroscopy, Opt. Express, 24(26), 30495–30504, doi:10.1364/OE.24.030495, 2016.

U.S. Energy Information Administration: Annual Energy Outlook 2015. [online] Available from: www.eia.gov/forecasts/aeo, 2015.

De Visscher, A.: Air Dispersion Modeling: Foundations and Applications, John Wiley & Sons, New York., 2013.

Wagner, G. A. and Plusquellic, D. F.: Ground-based, integrated path differential absorption LIDAR measurement of CO2 , CH4 , and H2O near 1.6 μm, Appl. Opt., 55(23), 6292–6310, doi:10.1364/AO.55.006292 1., 2016.

Waxman, E. M., Cossel, K. C., Truong, G.-W., Giorgetta, F. R., Swann, W. C., Coburn, S., Wright, R. J., Rieker, G. B., Coddington, I. and Newbury, N. R.: Intercomparison of Open-Path Trace Gas Measurements with Two Dual Frequency Comb Spectrometers, Atmos. Meas. Tech. Discuss., 1(March), 1–26, doi:10.5194/amt-2017-62, 2017.

Weil, J. C., Sullivan, P. P., Patton, E. G. and Moeng, C. H.: Statistical Variability of Dispersion in the Convective Boundary Layer: Ensembles of Simulations and Observations, Boundary-Layer Meteorol., 145(1), 185–210, doi:10.1007/s10546-012-9704-y, 2012.

Wilson, J. D. and Sawford, B. L.: Review of Lagrangian stochastic models for trajectories in the turbulent atmosphere, Boundary-Layer Meteorol., 78(1–2), 191–210, doi:10.1007/BF00122492, 1996.

Wu, C. F., Wu, T. gang, Hashmonay, R. A., Chang, S. Y., Wu, Y. S., Chao, C. P., Hsu, C. P., Chase, M. J. and Kagann, R. H.: Measurement of fugitive volatile organic compound emissions from a petrochemical tank farm using open-path Fourier transform infrared spectrometry, Atmos. Environ., 82(17), 335–342, doi:10.1016/j.atmosenv.2013.10.036, 2014.

Zavala-Araiza, D., Lyon, D. R., Alvarez, R. A., Davis, K. J., Harriss, R., Herndon, S. C., Karion, A., Kort, E. A., Lamb, B. K., Lan, X., Marchese, A. J., Pacala, S. W., Robinson, A. L., Shepson, P. B., Sweeney, C., Talbot, R., Townsend-Small, A., Yacovitch, T. I., Zimmerle, D. J. and Hamburg, S. P.: Reconciling divergent estimates of oil and gas methane emissions, Proc. Natl. Acad. Sci., 112(51), 15597–15602, doi:10.1073/pnas.1522126112, 2015a.

Zavala-Araiza, D., Lyon, D., Alvarez, R. A., Palacios, V., Harriss, R., Lan, X., Talbot, R. and Hamburg, S. P.: Toward a

Functional Definition of Methane Super-Emitters: Application to Natural Gas Production Sites, Environ. Sci. Technol., 49, 8167–8174, doi:10.1021/acs.est.5b00133, 2015b.

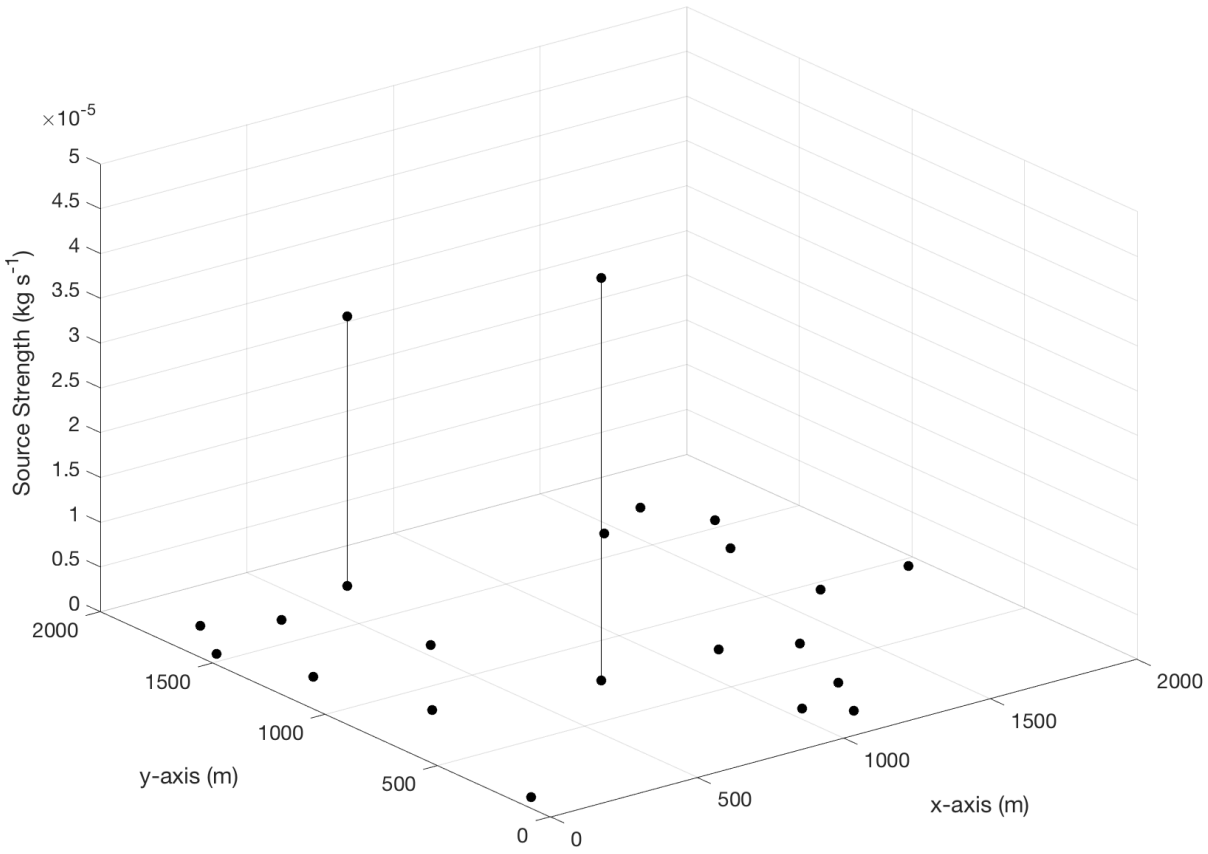

Figure 1: Synthetic test observation area: 2 km x 2 km domain with 20 source locations (black dots) at randomly distributed x and y locations (position shown on x and y axes). Of 20 point sources, well site 6 (x=750, y=750) and well site 19 (x=650, y=1750) have non-zero source strengths (shown on the z-axis).

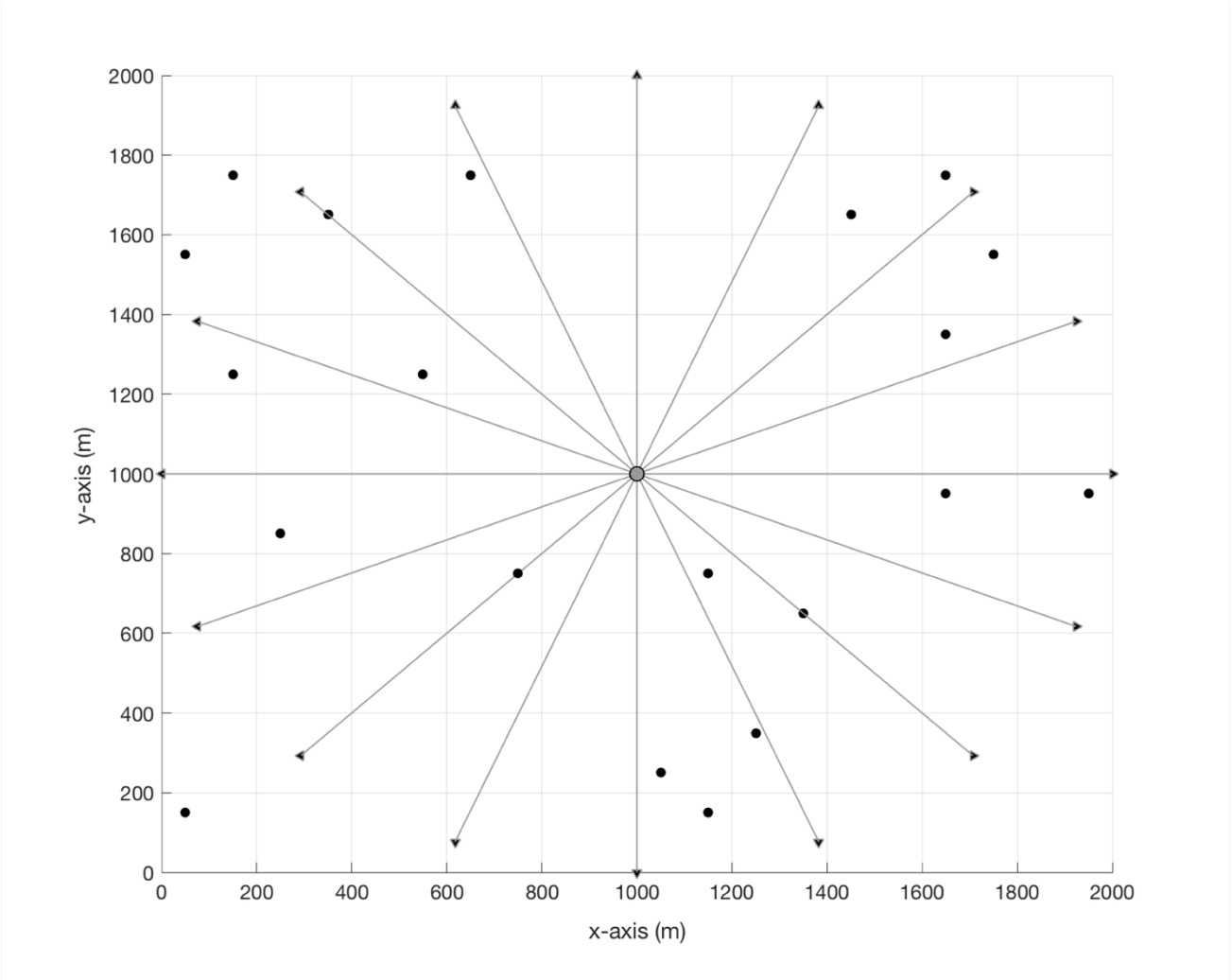

Figure 2: Map view of synthetic tests, with 20 source locations shown as black dots and 16 beams shown as gray lines that extend from the spectrometer (circle at x = 1000 m and y = 1000 m) to retroreflectors (black triangles).

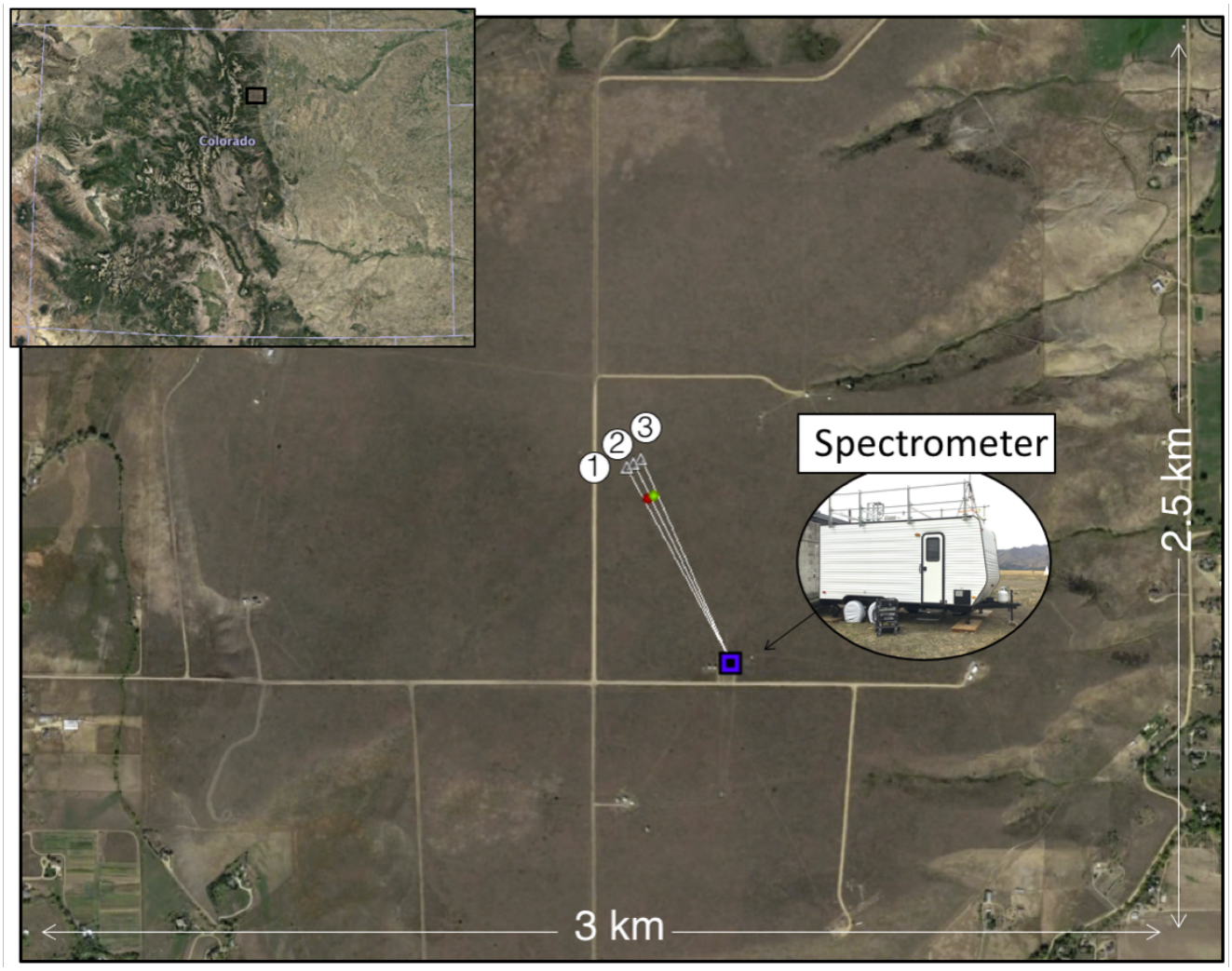

Figure 3: Map view of observation test site at Table Mountain, Colorado (upper left inset shows geographic location of test site), with two source locations (location 1, in red, between beams 1 and 2; location 2, in green, between beams 2 and 3) and three beams shown as white lines that extend from the spectrometer (blue square) to retroreflectors (white triangles, and labeled 1-3).

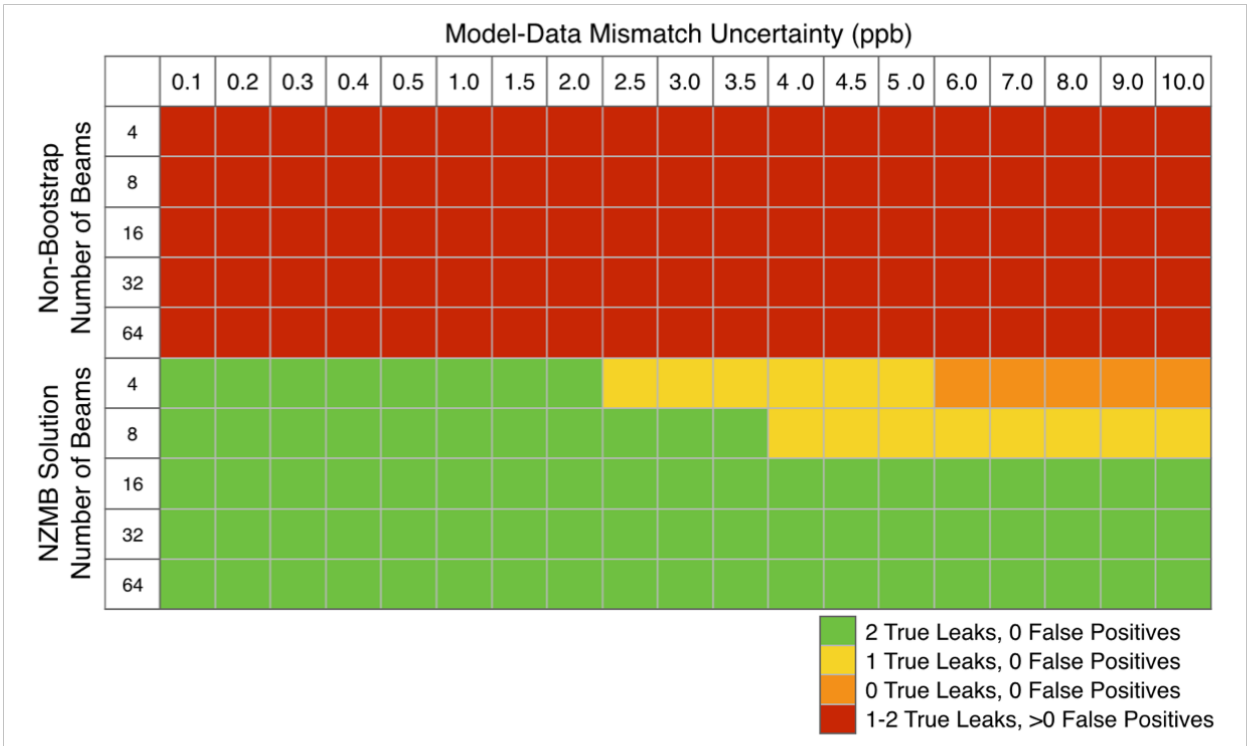

Figure 4: Summary of synthetic data test results. Top 5 rows show results of non-bootstrap inversions and bottom 5 rows show results of NZMB inversions for the 4, 8, 16, 32, and 64 beam cases. Columns indicate results for different values of model-data mismatch added as noise to the synthetic measurements. Color coding of cells indicates summary of model success, as detailed by the legend.

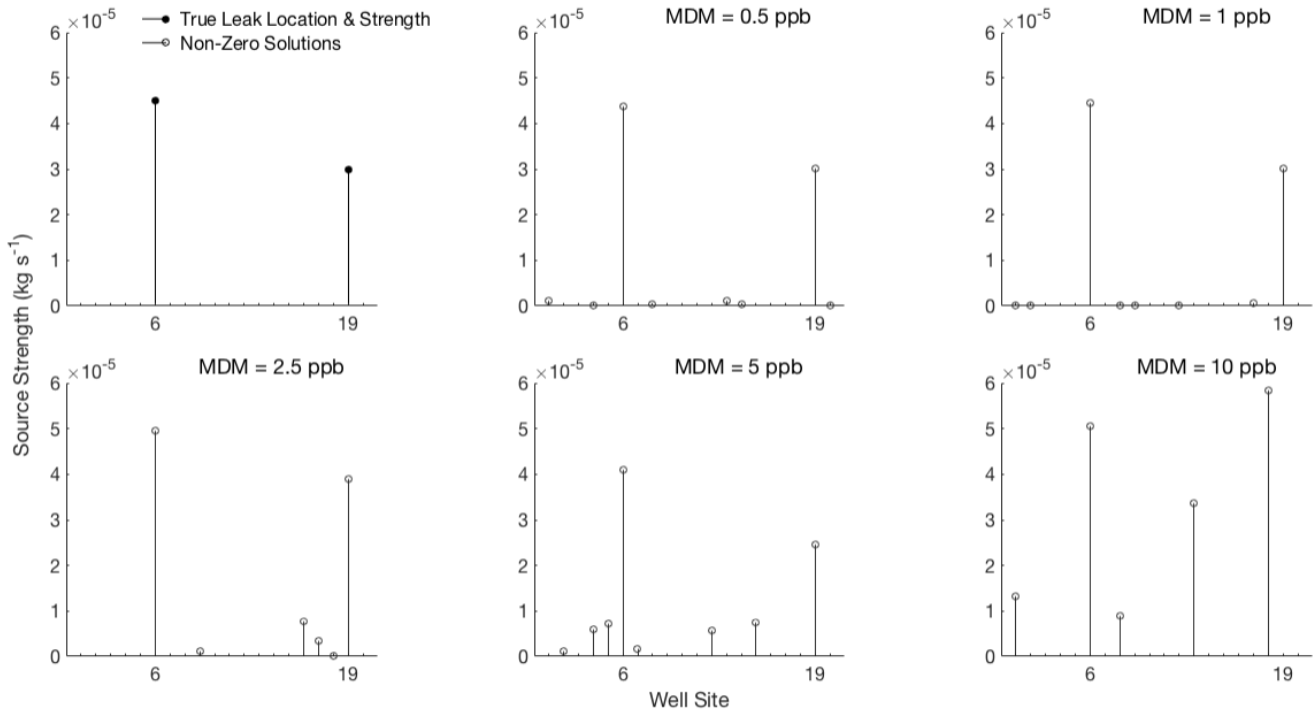

Figure 5: Top left panel shows well site numbers (x-axis) and corresponding "true" leak rates (y-axis), and remaining panels show resulting leak rate (y-axis) at each well site (x-axis) from non-bootstrap least squares fit to synthetic observations perturbed with model-data mismatch (MDM) noise shown, for the 8-beam case. Open circles show locations and strengths of all non-zero solutions.

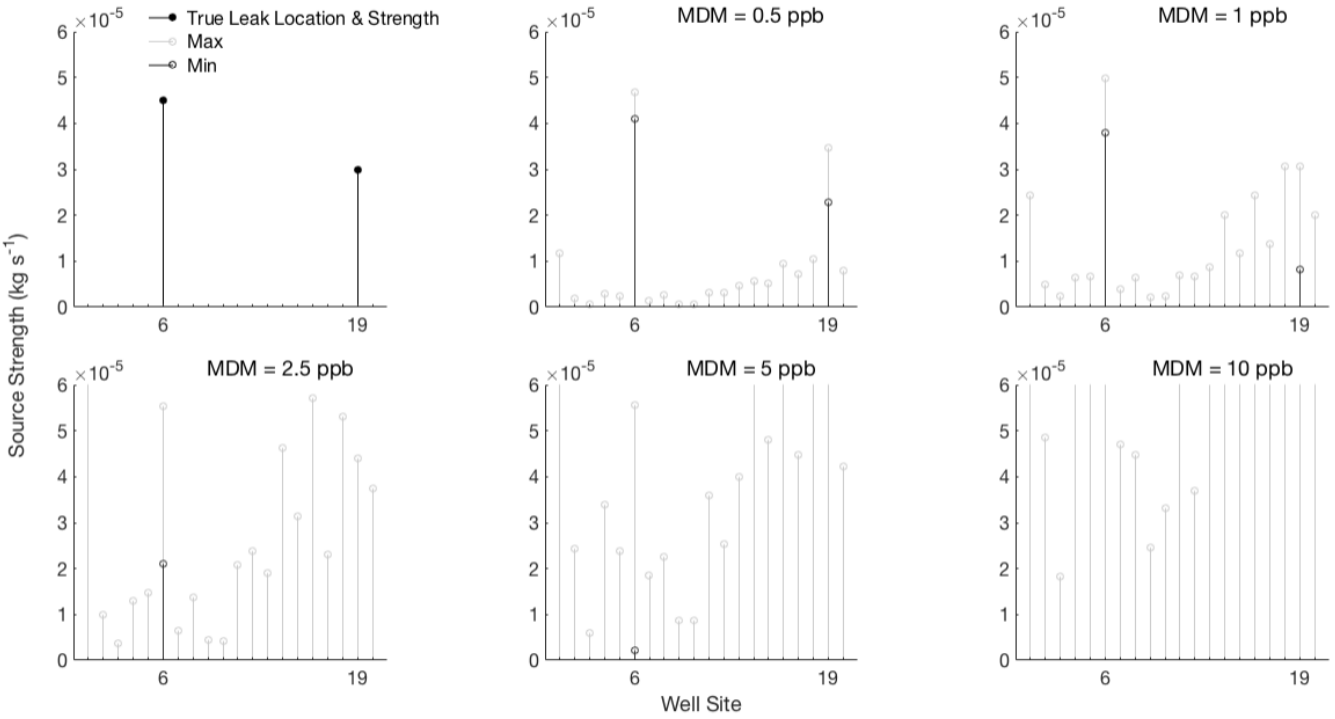

Figure 6: Top left panel shows well site numbers (x-axis) and corresponding "true" leak rates (y-axis), and remaining panels show NZMB results (y-axis) for each well site location (x-axis) with synthetic observations perturbed with model-data mismatch (MDM) noise shown, for the 8-beam case. Light gray (black) open circles show locations and strengths of the maximum (minimum) of 1000 bootstrap operations. Minimum values of zero are not plotted.

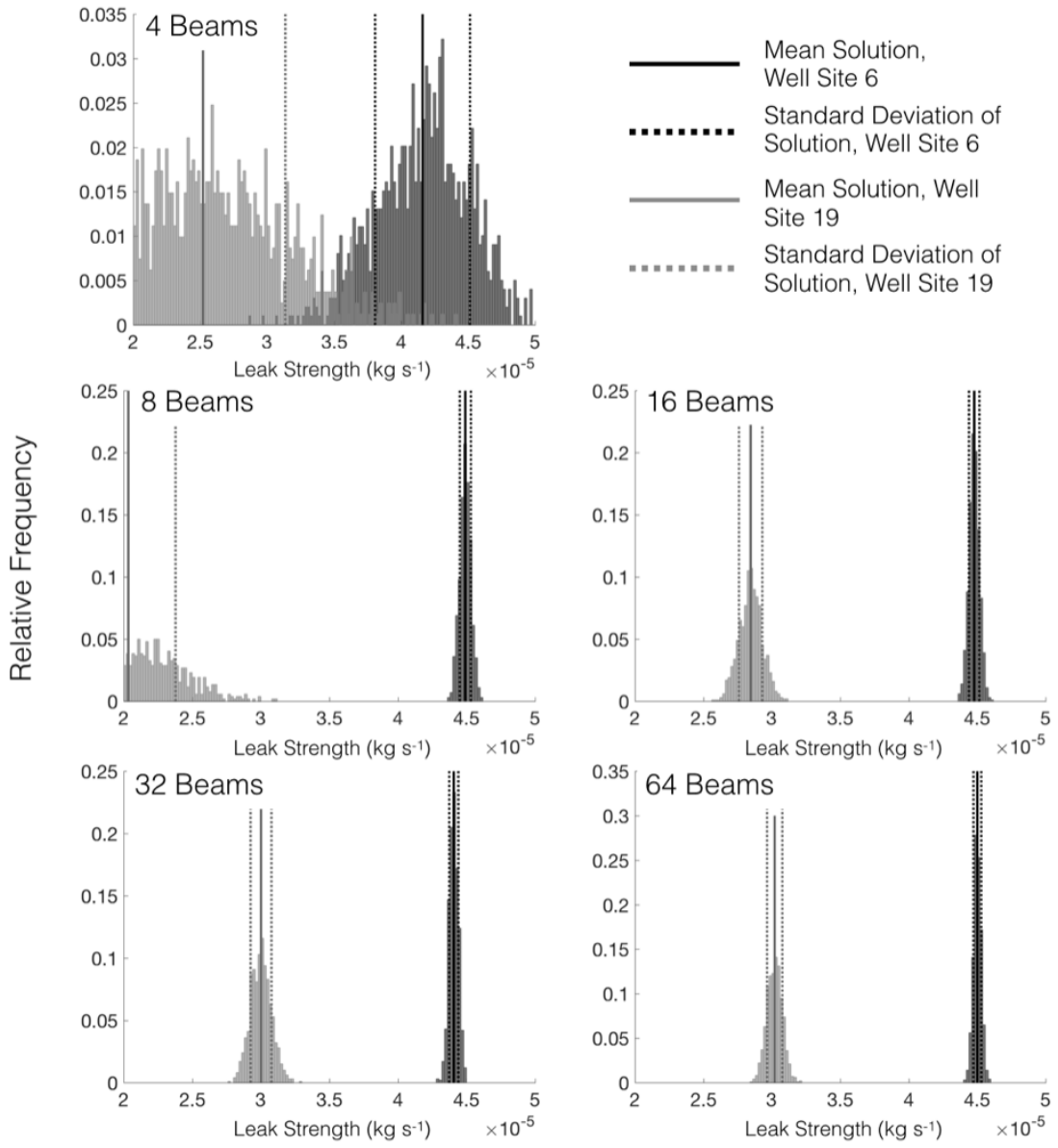

Figure 7: Histograms of source strength, with mean ± 1 standard deviation shown with vertical lines for well site 6 (black) and well site 19 (gray), for each beam configuration, and with 2 ppb model-data mismatch uncertainty. Note that x-axes are truncated at 2E-5 kg s^{-1} (lower bound) and 5E-5 kg s^{-1} (lower bound) for scale.

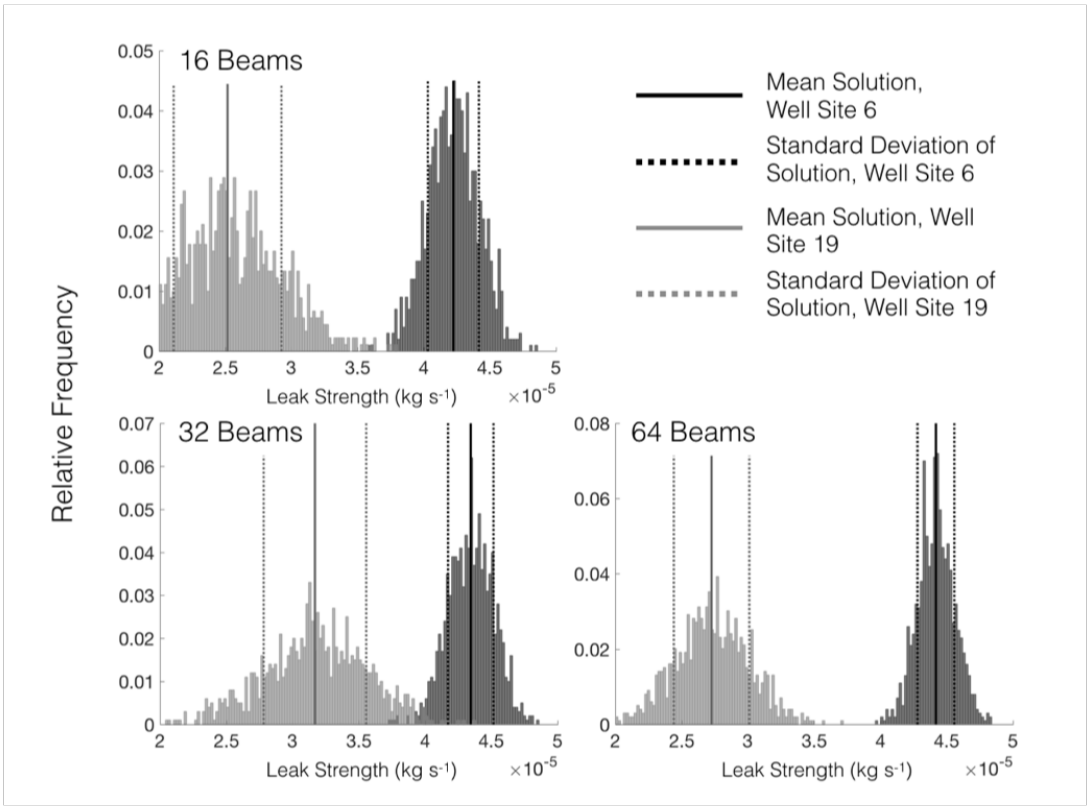

Figure 8: Histograms of source strength, with mean ± 1 standard deviation shown with vertical lines for well site 6 (black) and well site 19 (gray), for 16, 32 and 64 beam configurations, and with 10 ppb model-data mismatch uncertainty.

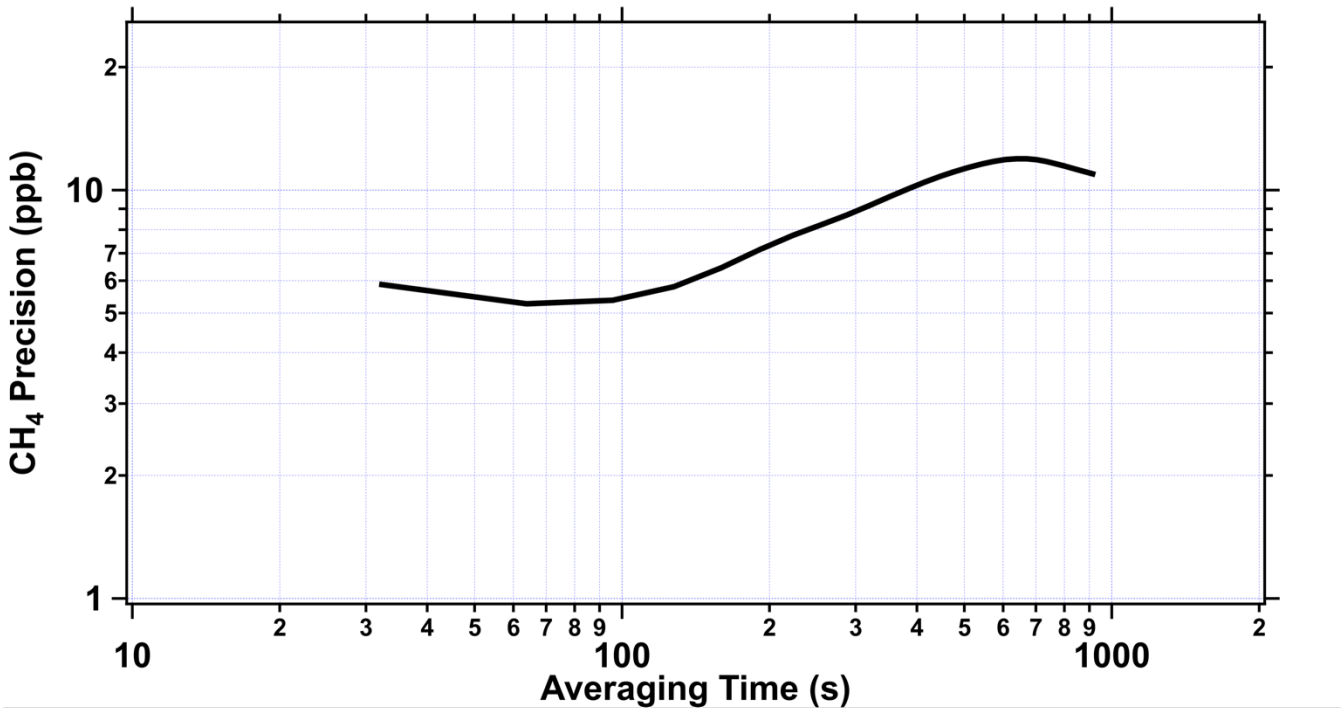

Figure 9: Allan deviation plot showing changes in measurement precision with averaging time from field data collected at Table Mountain on January 26th, 2017.

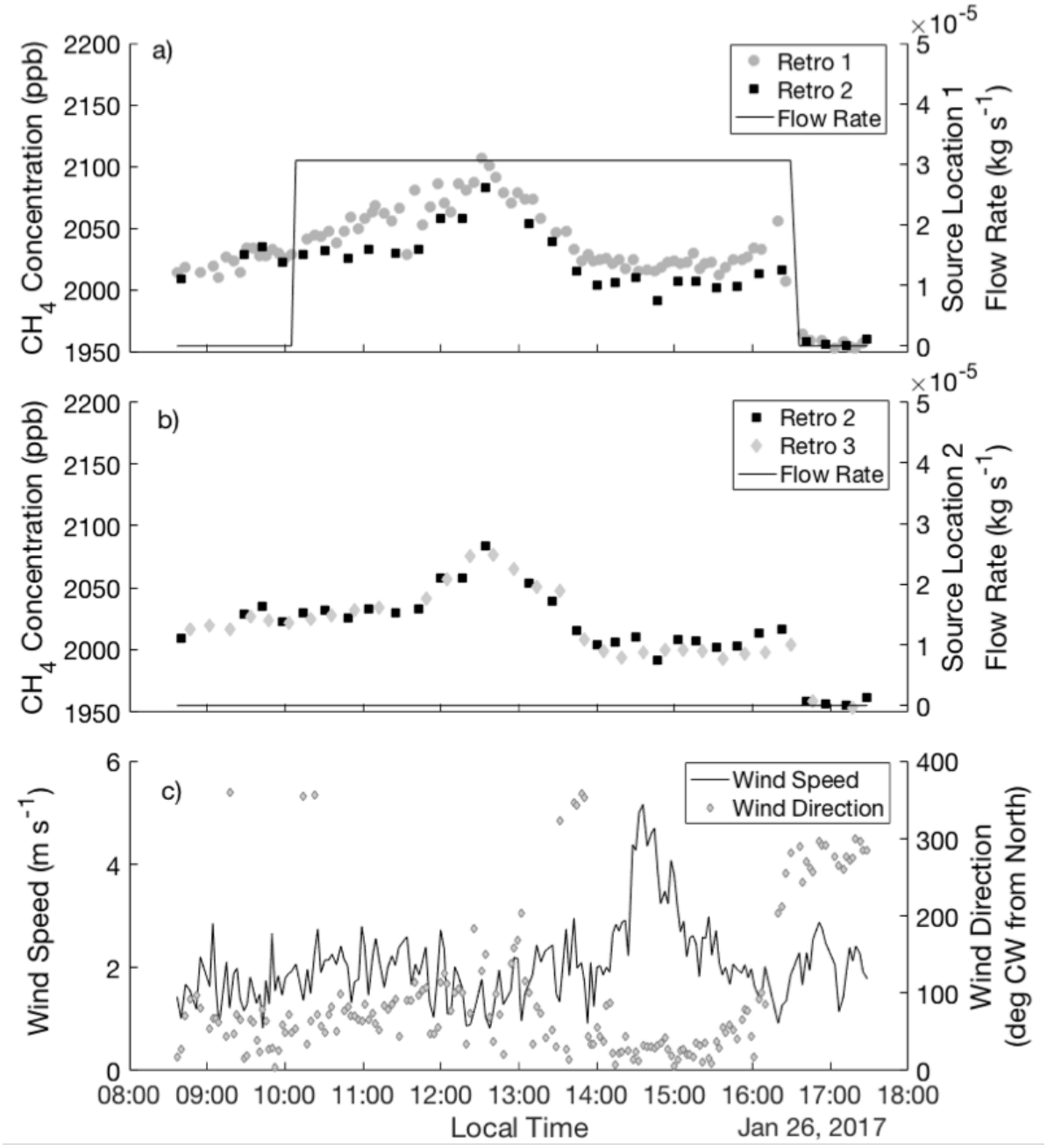

Figure 10: Line integrated atmospheric CH_4 concentrations measured on January 26, 2017 along beam paths to retroreflectors 1 and 2 (a), and to retroreflectors 2 and 3 (b), and wind speed and wind direction (c). ~~Each point represents a 2-minute averaged concentration.~~ Gray and black points and left-hand axes of panels (a) and (b) show CH_4 concentration. The black line and right-hand axis in panel (a) shows the flow rate at source location 1 (bounded by retroreflectors 1 and 2) and the black line and right-hand axis in panel b shows the flow rate at source location 2 (bounded by retroreflectors 2 and 3). In panel (c), the black line and left-hand axis show wind speed and the gray diamonds and right-hand axis show wind direction (according to meteorological convention, 0° = north, 90° = east, 180° = south, 270° = west, and 360° = north)~~speed~~. All data reflects 2-minute averaging time.

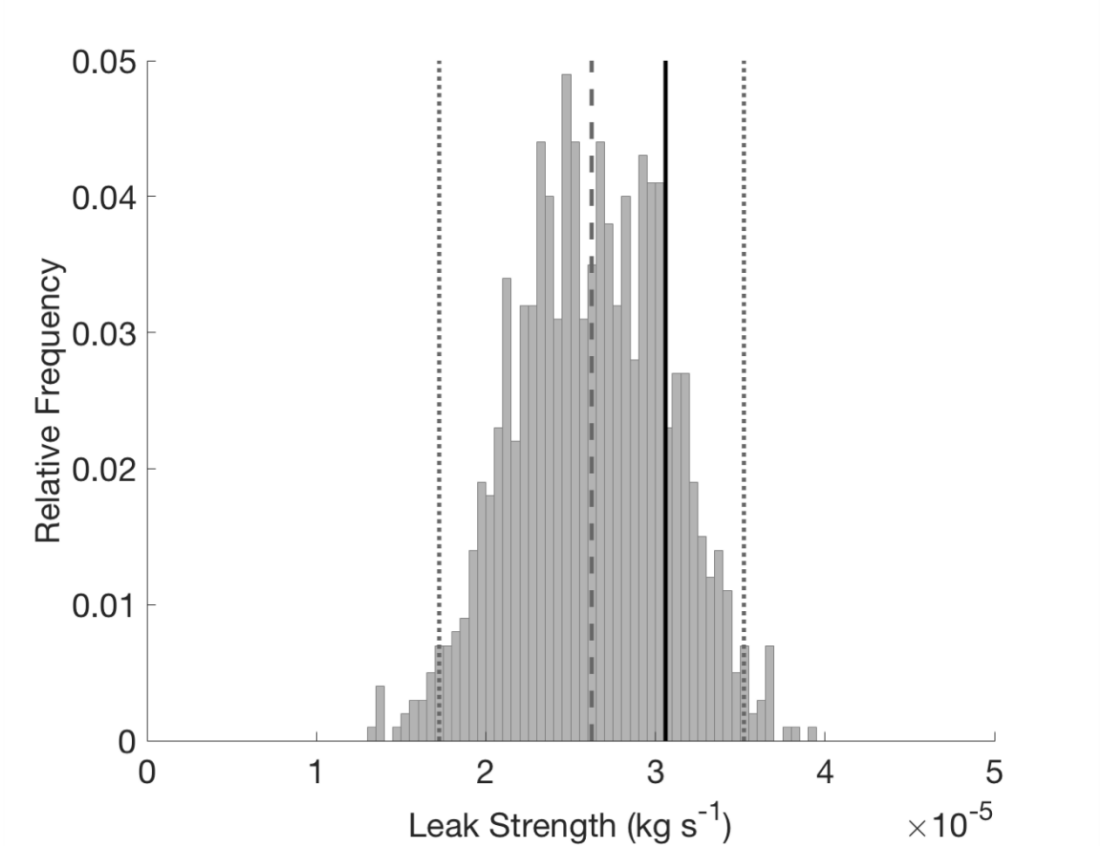

Figure 11: Histogram of NZMB estimated source strength at source location 1, with dashed line showing the bootstrap mean and thin dotted lines showing ± ~~1~~ 2 standard deviation. The thick black line shows the true leak strength at source location 1 (3~~2~~.1 E-5 kg s^{-1}).

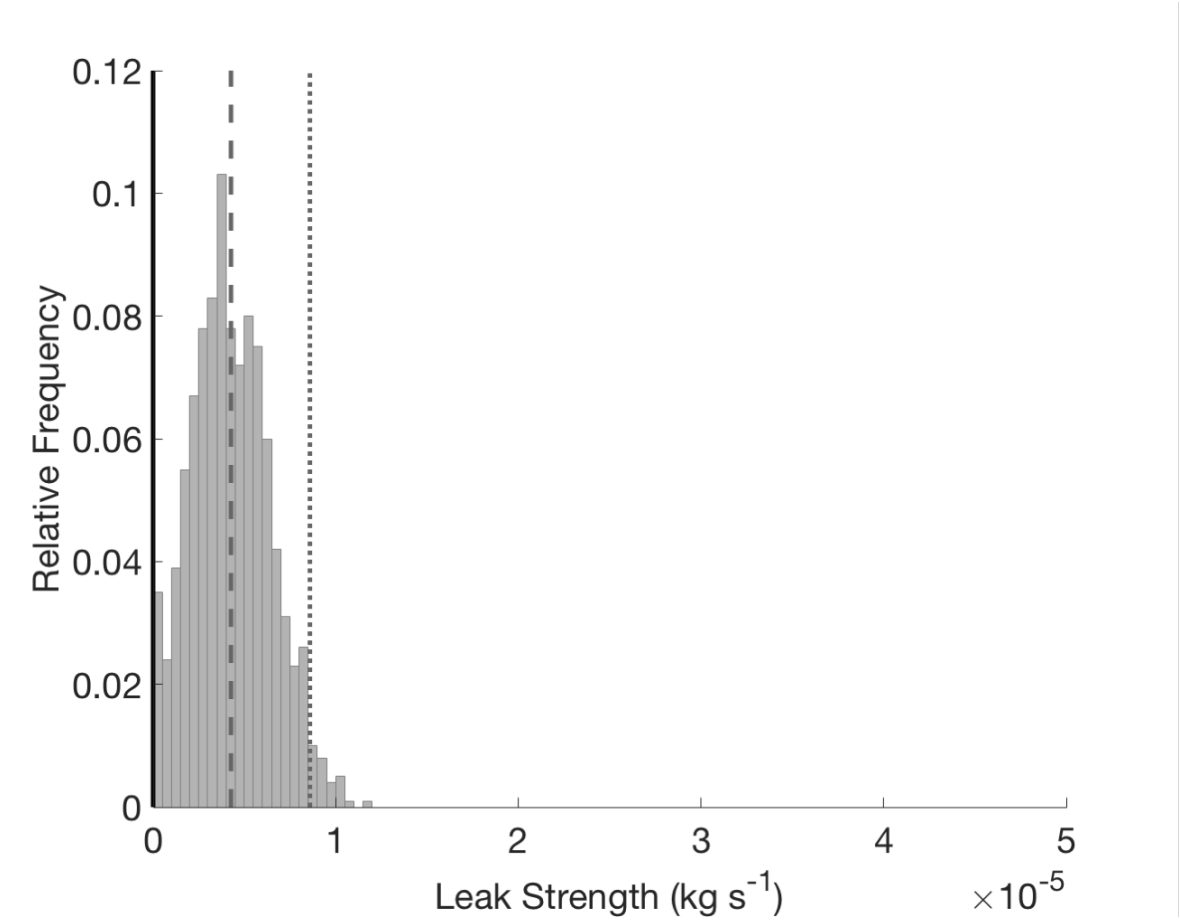

Figure 12: Histogram of NZMB estimated source strength at source location 2, with dashed line showing the bootstrap mean and thin dotted lines showing ± ~~1~~ 2 standard deviation. The thick black line shows the true leak strength at source location 2 (0 kg s^{-1}). The presence of 0 kg s^{-1} in the histogram triggers acceptance of the null hypothesis (that the emissions rate at this site is zero).

Number of Beams	Well Site 6 Mean Strength	Leak One 1 s. d.	Well Site 19 Mean Strength	Leak Two 1 s. d.
4	4.2E-5 kg s^{-1}	0.4E-5 kg s^{-1}	2.5E-5 kg s^{-1}	0.6E-5 kg s^{-1}
8	4.5E-5 kg s^{-1}	0.4E-6 kg s^{-1}	2.0E-5 kg s^{-1}	0.3E-5 kg s^{-1}
16	4.5E-5 kg s^{-1}	0.4E-6 kg s^{-1}	2.8E-5 kg s^{-1}	0.9E-6 kg s^{-1}
32	4.4E-5 kg s^{-1}	0.3E-6 kg s^{-1}	3.0E-5 kg s^{-1}	0.8E-6 kg s^{-1}
64	4.5E-5 kg s^{-1}	0.3E-6 kg s^{-1}	3.0E-5 kg s^{-1}	0.6E-6 kg s^{-1}
True Leak: 4.5E-5 kg s^{-1}			**True Leak: 3.0E-5** kg s^{-1}	

Table 1: NZMB Solutions for leak strength of true leaks, given 2 ppb model-data mismatch uncertainty, for each beam configuration.

	Source Location 1	Source Location 2
Controlled Leak Time On:	10:08	NA
Controlled Leak Time Off:	16:30	NA
Measured Mean Flow Rate:	32.1 E-5 ± 0.01E-5 kg s^{-1}	0.0 ± 0.0 kg s^{-1}
Non-Bootstrap Solution:	2.43 E-5 kg s^{-1}	00.52 E-5 kg s^{-1}
NZMB Solution:	2.60 E-5 ± 0.52 E-5 k/s	0.0 ± 0.0 kg s^{-1}

Table 2: Controlled methane release flow rates and 1 standard deviation for each field experiment, including local time that leak was turned on and off.

Supplemental Information for: Methane leak detection and sizing over long distances using dual frequency comb laser spectroscopy and a bootstrap inversion technique

Caroline B. Alden[1,2*], Subhomoy Ghosh[3], Sean Coburn[1], Colm Sweeney[2,4], Anna Karion[3], Robert Wright[1], Ian Coddington[3], Kuldeep Prasad[3], Gregory B. Rieker[1]

[1]Department of Mechanical Engineering, University of Colorado at Boulder, Boulder, CO, 80309 USA
[2]Cooperative Institute for Research in Environmental Sciences, Boulder, CO 80309, USA
[3]National Institute of Standards and Technology (NIST), Gaithersburg, MD, 20899, USA
[4]National Oceanic & Atmospheric Administration (NOAA), Boulder, CO 80305, USA

Correspondence to: Caroline B. Alden (caroline.alden@colorado.edu)

Pasquill-Gifford Atmospheric Stability Class

Stability classes for each time period of the field data collected on 1/26/2017 are determined following Turner's Method (Turner, 1964), as outlined in section 6.4.1 of the U.S. EPA's Meteorological Monitoring Guidance for Regulatory Modeling Applications Meteorological Monitoring Guidance Table 6-5 (Bailey, 2000). For each measurement time, we follow the procedure outlined in Table 6-6 of that report to determine the Net Radiation Index number. The first step is to calculate solar altitude and use Table 6-5 in that report to determine the insolation class number. The second step is to modify the insolation number based on total cloud cover, which we estimate based on the difference between expected maximum and actual downwelling solar radiation as measured at Table Mountain (our field site) by the National Oceanic and Amtospheric Administration (NOAA) Earth System Research Laboratory (ESRL) Surface Radiation Network (SURFRAD). That data is available at the following FTP site ftp://aftp.cmdl.noaa.gov/data/radiation/surfrad/Boulder_CO/. Further information specific to the SURFRAD site located at our field site on Table Mountain can be accessed here: https://www.esrl.noaa.gov/gmd/grad/surfrad/tablemt.html. We do not have information about the ceiling at Table Mountain, so modifications of the insolation class number for ceiling height are not made. The third step is to determine the stability class for each time period using the derived Net Radiation Index number and measured wind speed in Table 6-4.

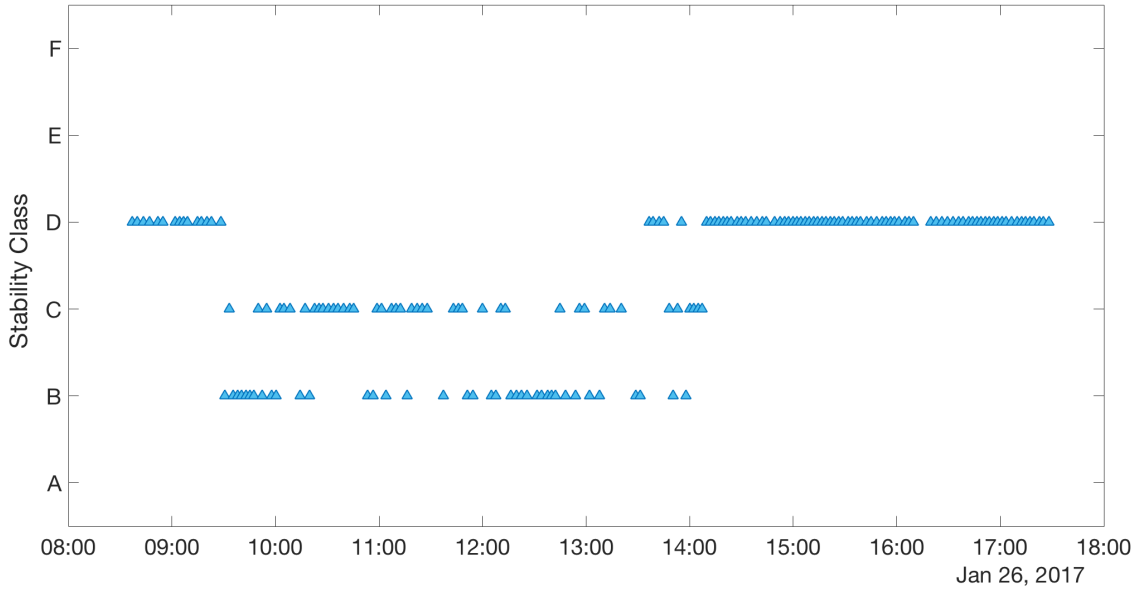

Figure S1: Timeseries of stability class calculated for Table Mountain field site on 01/26/2017.

As a sensitivity test, we examine the use of dispersion coefficients that are based on 10-minute empirical data for our 2-minute datasets. We calculate the expected potential error introduced in σ_y using the standard equation for the adjustment of dispersion parameters for different averaging times:

$$\sigma_{y,2} = \sigma_{y,1}\left(\frac{t_2}{t_1}\right)^p, \quad \text{(S1)}$$

found in Gifford (1976).

The Briggs parameters used in this study were calculated using an averaging time of 10 minutes, so that t_1 = 10 minutes (Briggs, 1974; Griffiths, 1994). We use measurements averaged over 2 minutes and meteorological data averaged over 2 minutes, so that t_2 = 2 minutes. Using the typical value of 0.2 for the empirical parameter p (for t_2 < 1 hour) (Gifford, 1976), we calculate that a factor of 0.7 correction should be applied to the Briggs horizontal dispersion coefficient, σ_y. We are not aware of similar adjustment formulas for the vertical dispersion coefficient.

We apply the correction factor to the field data collected on January 26th, 2017 (the synthetic data has no time dimension, and transport is considered perfect in that observing system simulation experiment, so the choice of the dispersion coefficient has less bearing on the findings). The results are seen in Table S1.

	Source Location 1	**Source Location 2**
Controlled Leak Time On:	10:08	NA
Controlled Leak Time Off:	16:30	NA
Measured Mean Flow Rate:	3.1 E-5 ± 0.01E-5 kg s^{-1}	0.0 ± 0.0 kg s^{-1}
Non-Bootstrap Solution:	2.4 E-5 kg s^{-1}	0.5 E-5 kg s^{-1}
NZMB Solution:	2.6 E-5 ± 0.5 E-5 k/s	0.0 ± 0.0 kg s^{-1}

Table S1: Same as Table 2 in the main text, but calculated with an adjustment to the dispersion coefficient, σ_y. Controlled methane release flow rates and 1 standard deviation for each field experiment, including local time that leak was turned on and off.

The recovered fluxes do not change to within a detectable range. It is possible that the sensitivity of our result to the adjustment to σ_y is low because the cross-wind fluctuations are largely averaged through by the laser beam when winds are orthogonal to the beam. The histograms of the solutions shown in Table S1 are shown below.

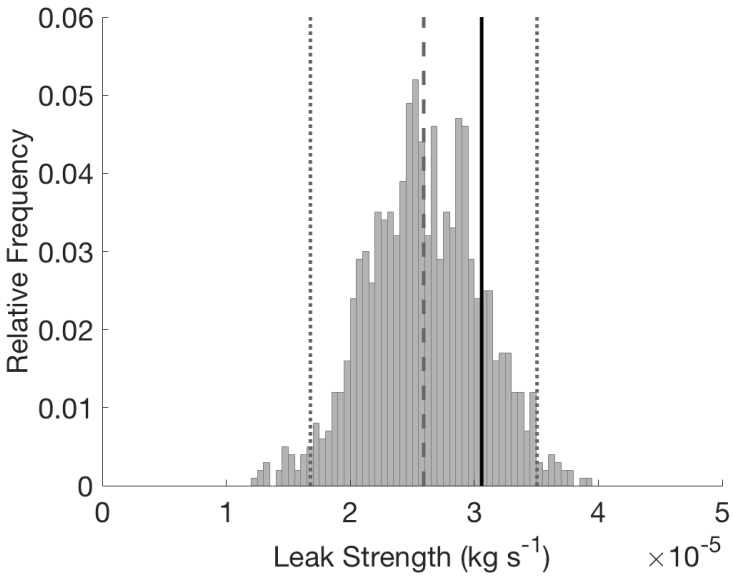

Figure S2: Same as Figure 11 in the main text, but calculated with an adjustment to the dispersion coefficient, σ_y. Histogram of NZMB estimated source strength at source location 1, with dashed line showing the bootstrap mean and thin dotted lines showing ± 2 standard deviation. The thick black line shows the true leak strength at source location 1 (3.1 E-5 kg s^{-1}).

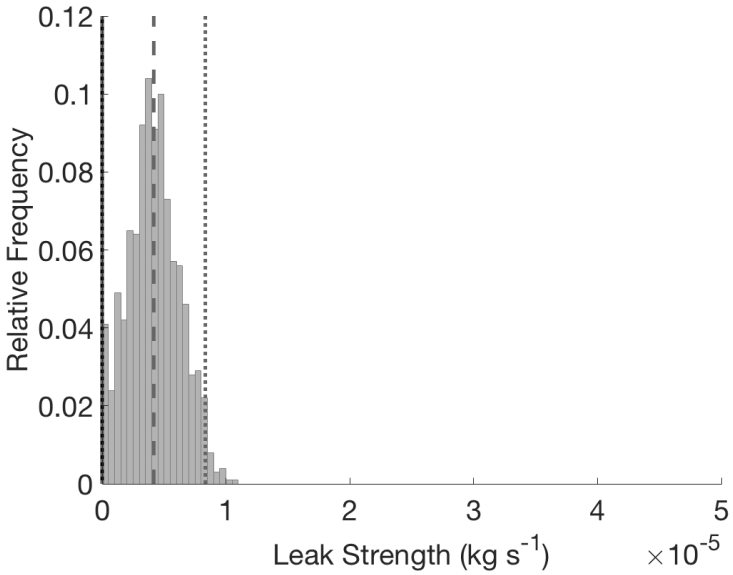

Figure S3: Same as Figure 12 in the main text, but calculated with an adjustment to the dispersion coefficient, σ_y. Histogram of NZMB estimated source strength at source location 2, with dashed line showing the bootstrap mean and thin dotted lines showing ± 2 standard deviation. The thick black line shows the true leak strength at source location 2 (0 kg s^{-1}). The presence of 0 kg s^{-1} in the histogram triggers acceptance of the null hypothesis (that the emissions rate at this site is zero).

Full range of Model-Data Mismatch Results

Here we show the full range of model-data mismatch results, which are not shown for the sake of avoiding figure complexity in Figures 5 and 6 in the main manuscript.

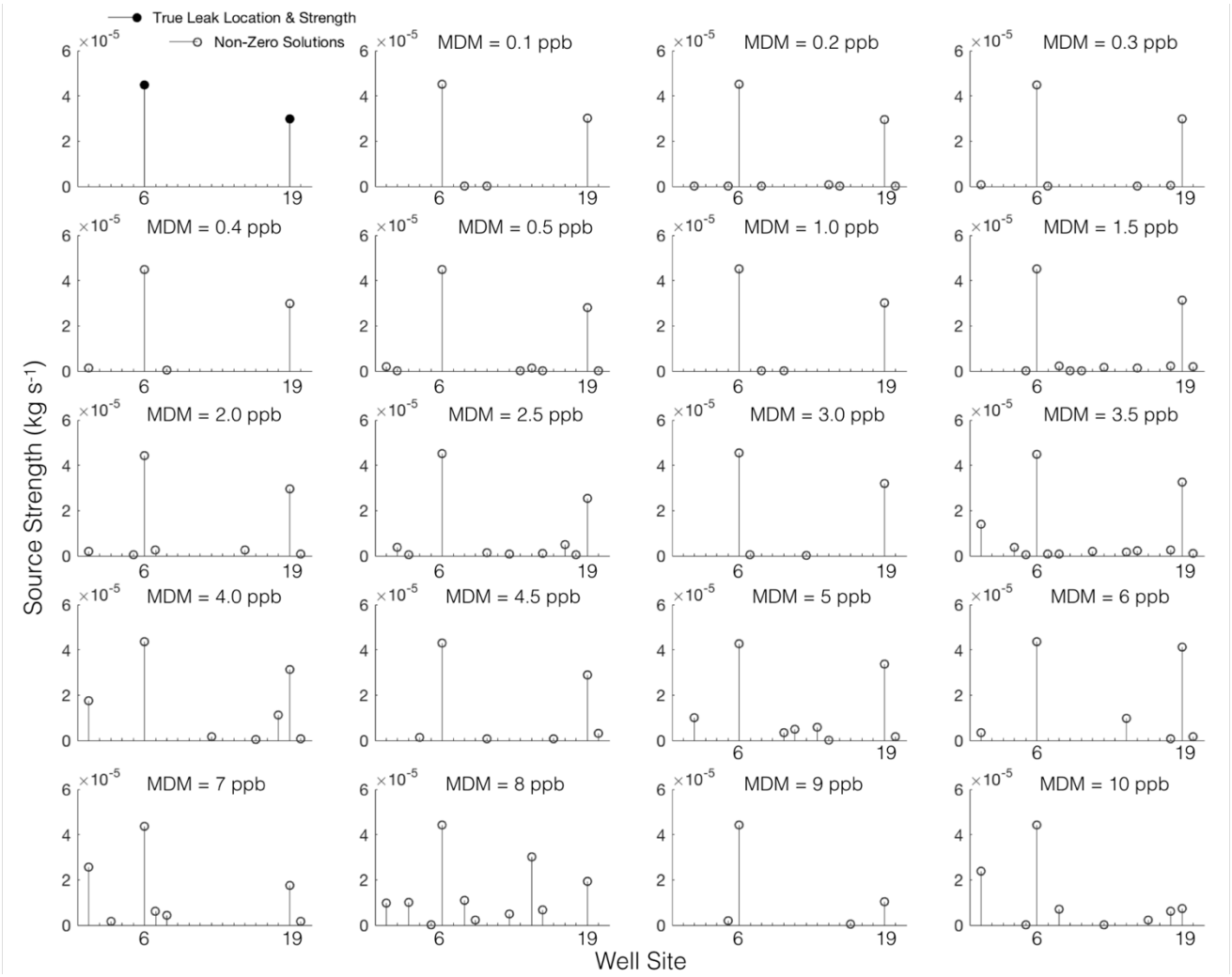

Figure S4: Same as main text Figure 5, but with all model-data mismatch cases shown.

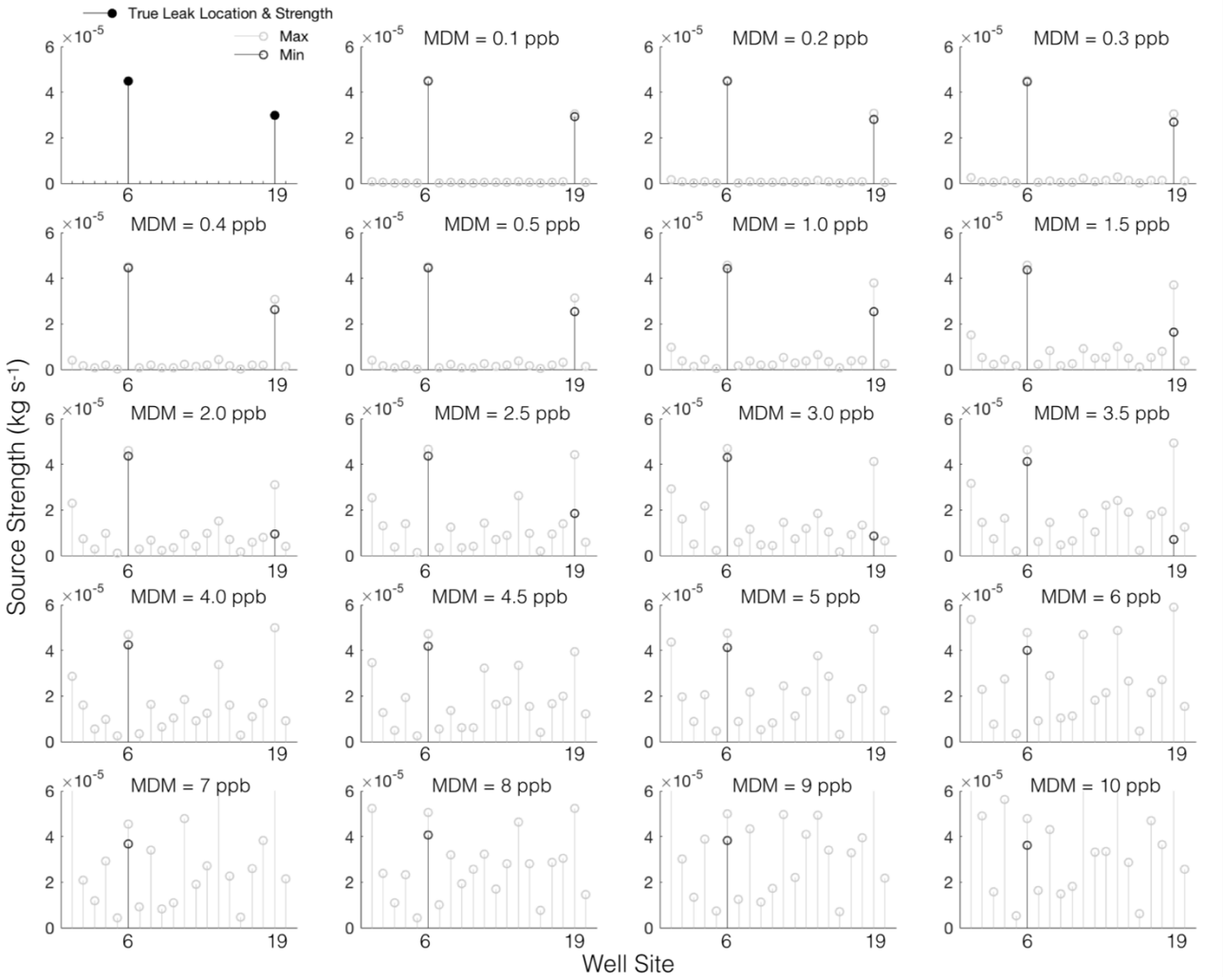

Figure S5: Same as main text Figure 6, but with all model-data mismatch cases shown.

Estimation of transport uncertainty

A simple estimate of transport uncertainty is derived by considering the combined impacts on simulated atmospheric concentrations of meteorological measurement instrument uncertainties, parameterization of atmospheric stability, and placement of the sonic anemometer relative to leak location (that is, the influence of using a point measurement of wind speed to characterize the entire wind field over the Table Mountain area). We use meteorological measurements made by two separate sonic anemometers placed at different locations (several hundred meters apart) on the Table Mountain site over the course of nearly 6 hours on March 3, 2017. Wind speed and wind direction are measured by both instruments. These meteorological measurements are used in a plume model simulation of the enhancement due to a leak with a rate equal to the controlled release of methane at source location 1 (3.1E-5 kg s^{-1}) along a 585 m beam that is either 9 m (the mean minimum lateral distance between each leak and adjacent beam in our test configuration) or 209 m (the mean distance between each leak and each segment of the beams directly monitoring it) downwind of the leak and perpendicular to wind direction. We calculate the standard deviation in the enhancement along the beam given the simulations with data from different

anemometers and given an increase and a decrease in the stability of one stability class from the mean state calculated on January 26th, 2017 (e.g., for a daily mean stability class of B, the absolute differences between B and A and B and C are averaged). This standard deviation is then taken to be the 1-σ uncertainty due to transport reported in the comparison to model-data mismatch in Section 4.4.

Note on units

The units ppb are equal to the SI units nmol mol^{-1}.

References

NOAA Earth System Research Laboratory: Surface Radiation Budget (SURFRAD) Network Observations. Radiation data from subset "Boulder_CO". NOAA National Centers for Environmental Information. Accessed 2017.

Bailey, D. T.: Meteorological Monitoring Guidance for Regulatory Modeling Applications Meteorological Monitoring Guidance, Report No. EPA-454/R-99-005., 2000.

Briggs, G. A.: Diffusion estimation for small emissions, in ATDL Contribution File No. 79, Air Resources Atmospheric Turbulence and Diffusion Laboratory, NOAA, Oak Ridge, Tennessee., 1974.

Gifford, F. A.: Atmospheric dispersion models for environmental pollution, in Lectures on Air Pollution and Environmental Impact Analysis, edited by D. A. Haigen, pp. 35–58, Boston, Mass., 1976.

Griffiths, R. F.: Errors in the use of the Briggs parameterization for atmospheric dispersion coefficients, Atmos. Environ., 28(17), 2861–2865, doi:doi:10.1016/1352-2310(94)90086-8, 1994.

Turner, D. B.: A diffusion model for an urban area, J. Appl. Meteorol., 3, 83–91, 1964.