In general, we agree with many of the reviewer's observations about the difficulties in developing realistic environments that truthfully represents the dynamics of operational oil and gas facilities. It is the authors' experience that complex and unique hard surfaces do contribute to this dynamic turbulent environment but are not the sole, and maybe not even the prime, contributor to uncertainties in such an environment. Many operational oil and gas facilities span vast areas and are surrounded by other similar facilities. These and other natural and man-made dynamic activities on or near a site of interest also play significant roles in both near- and far-field observations of methane releases. The influence of this collective set of items along with one's a priori knowledge about where leaks might or might not happen all contribute to the dynamic observation environment, and none should be considered, at this point in time, the prime or dominant contributor.

Our manuscript acknowledges the existence of this complexity, identifies the shortcomings of the chosen test facility, and presents the results using a reference protocol as a framework for conducting and reporting these activities. We have responded to your comments and have addressed those that we believe fall within the scope of this paper.

- Generally, this manuscript's title, abstract and introduction are a bit misleading. There are multiple references to the METEC test facility, despite these tests not being performed at the METEC or even at a similar facility. These blind tests were performed in an open field without flow obstructions or realistic emitting equipment.

  All references to METEC are either to credit them with development of the continuous monitoring protocol (CMP), to reference other testing that has been performed there, to contrast our testing methodology with that of METEC, or in reference to the METEC representative that advised and audited our testing. Nothing in our manuscript implies that our testing was performed at "METEC or even a similar facility." We explicitly state that our testing was NOT performed at METEC and that our site had no oil and gas equipment.

  The CMP was established such that it is separable from the METEC facility and may be extended to other test environments as a common framework for assessing a disparate set of technologies for detecting and quantifying CH4.

  The language in the abstract was adjusted to remove the potential interpretation that *System performance was evaluated against metrics… "at"*

Colorado State University, and instead identify performance was *"... evaluated against metrics defined in the Continuous Monitoring Protocol established by the Colorado State University Methane Emissions Technology Evaluation Center (METEC)."*

- These tests do not appear to significantly differ from other similar papers describing the dual laser-head tomographic approach.

  Please cite the papers you are referring to. Only two papers have been published regarding the use of this system to quantify methane emissions. One is about controlled releases at the Total TADI site, with known emission location, time, duration, and in most cases, rate. The other is about large-area fugitive emissions from an oil sands tailing pond and open-pit mine. Neither paper documents blind testing with controlled releases at unknown times, locations, and rates.

- There is no oil and gas equipment present and so references to other METEC studies or the implications of these results for the oil and gas field in general should probably be kept to the conclusions or future work sections. For example, It is not reasonable to make assertions of "localization" to equipment units or groups when there was no equipment present.

  Use of the unit/group terminology was done to convey relative scale in the localization accuracy of emissions because that is the terminology used in the CMP. For additional clarity, we have changed "Equipment Unit" to "Cell Unit" and "Equipment Group" to "Cell Group". Again, the intent of the CMP developed by METEC was to establish a common test methodology that could be transported to *any* test location capable of conducting controlled releases, with or without the presence of actual or mock production equipment. The study team included Bell from CSU METEC to audit the test setup and confirm the approach was as similar and comparable to the METEC implementation as possible, while enabling the GreenLITE technology to demonstrate its capability at a more realistic scale deployment.

- The concept of "blind" testing of rates (quantification) should be removed from title, given the rates were not technically blinded (see below).

  A small number of non-blind tests were conducted to confirm functionality of the release system and to establish meaningful bounds on the release rates that should be tested in order to maximize the utility of the limited gas quantity available for this testing. One of our objectives was to establish the probability of detection curve, and it would not have been a prudent use of time or gas to perform releases at rates orders of magnitude below or above those needed to establish the PD curve.

The release rate range did not in any way limit the estimated emissions. In other words, if our estimate of emissions rate fell below the minimum possible release rate or above the maximum possible release rate, we did not adjust our estimate to fall within the bounds. For example, Release 05-01 was performed at 0.215 kg/h, but our estimated emission rate was only 0.072 kg/h. We did not force this up to 0.215 kg/h just because we did not expect a release at such a low rate.

The release rate, location, time, or duration was never provided to the observers/analysts, who did not even know on which days releases were to be performed.

- The concept of “continuous” should likely also be reconsidered in the title given that “frost heaving” required the system only be run for certain portions of the day. It was not clear what day-to-day manual corrections were made to correct for frost heaving? These should be described more fully.

The GreenLITE measurement was, has been, and continues to be deployed in an autonomous and continuous fashion. When possible, the system is installed such that movement is negligible over several months (e.g., posts set below frost line, reflectors mounted on buildings). The methods used in this study were designed to minimize interference with existing activities and infrastructure, which, in turn, placed additional constraints on the validated measurement periods. The system did run autonomously and continuously for nearly all of the 6-week testing window (with the exception of the events identified in section 3.4). The regular alignment adjustments were comparable to routine calibrations or comparisons to reference standards that are commonly employed by others and were only needed due to installation constraints.

From section 2.3:

“Regular realignment was performed via the system’s remote interface on all testing days, including both days when releases were performed and days when no releases were performed. Such realignment is not needed when GreenLITE™ is installed in a semi-permanent or permanent fashion, as was done in the deployments described by Dobler et al. (2017), Pernini et al. (2022) and Blakley et al. (2020).”

We added text to explain that “regular realignment” means making small adjustments to the programmed scanner positions that are used to point the sensor head at each of the reflectors to compensate for movement in the mounting structure.

- The introduction and abstract of this paper do not adequately convey that two different laser heads were used in a tomographic approach. This is an important aspect of the system. This makes the detailed comparisons to different open-path continuous monitoring systems irrelevant – they are fundamentally different approaches. Generally, it should be compared with the broad range of other continuous monitors (e.g., Bell et al. 2023).

  A single GreenLITE system was employed in this work. The reviewer is correct that this system comprises two functionally equivalent sensor systems. While a single GreenLITE sensor may be used to measure gas concentrations and estimate emissions, the full capability of GreenLITE to generate 2-D maps of concentration and emissions may only be realized when a pair of sensors (a system) is utilized.

  Text has been added to the introduction to clarify that two sensor units were used.

  Performance comparisons of fundamentally different approaches is the entire point of the CMP. However, a taxonomy of the many different approaches to continuous methane emission monitoring is outside the scope of this report. No detailed comparisons to different open-path continuous monitoring systems or point sensor networks were made. Instead, we refer the reader to the relevant publications such as Bell et al. 2023 for the performance results of those tested solutions, and only highlight key differences in the experimental design of this work and those studies.

- This paper seeks to draw distinctions between the tests performed in a field without realistic oil and gas equipment, with other tests performed at METEC. It is fine to reference and use the METEC protocols, but the detailed comparisons with METEC-published results in the introduction are misleading, given that these tests are fundamentally different on so many levels.

  The shortcomings of the site were provided in the manuscript. No comparison of results was made, only to experimental design. As the reviewer makes abundantly clear, the METEC facility is the most notable use of the protocol, and we feel it is appropriate to compare the application of the protocol at that facility to our application of the protocol, while clearly delineating the differences.

Also note that while METEC is recognized as a leading "realistic" testing facility, there are still many limitations of the METEC facility. For instance, METEC does not have operational equipment such as burners for process heating, or natural gas compressors, both of which introduce forced air currents and buoyant exhaust plumes at real operational facilities. Further, testing at the facility has not been conducted with operational emission sources, such as venting from pneumatic controllers, simulated through transient controlled releases. Therefore, the ONLY fundamental difference between the testing conducted here, and the testing conducted at METEC is the presence of small eddies and mixing in local wind fields around blunt body objects.

- Some critical differences between the tests performed here and tests performed at METEC:
  - In this test, no real oil and gas equipment was used, confounding source characteristics as well as turbulent mixing.

    This was an unfortunate drawback to our selected test site, which we address in section 2.2. We have added text about turbulence in this section to clarify why the lack of equipment makes our test site potentially less challenging than METEC. However, acknowledging this limitation, the results still demonstrate a) the flexibility/transportability of the METEC continuous monitoring protocol, and b) an initial performance assessment of the GreenLITE system.

  - In this test, high-purity gas was used, not a realistic gas mixture.

    We do not see how this is relevant. The CMP does not specify a gas mixture and states that gas composition is to be applied to the flow rate of the controlled releases to calculate the mass flow rate of each gas species. Use of a "realistic gas mixture" would have just added an additional calculation in the conversion of total gas flow rate to CH4 mass flow rate. The presence of other gas species has no effect on the GreenLITE measurement. Knowledge of the gas composition is what matters, not the composition itself.

    Additionally, testing at METEC utilizes compressed natural gas (CNG) from the local distribution system. CNG is also a higher purity gas than would be found at most upstream oil and gas facilities, as it has been processed to meet pipeline specifications including heating value, moisture content, and limitations on the presence H2S or other toxic species.

- In this test, guidance was provided to testers beforehand about what rates the releases should happen at; this does not occur at METEC.

  Pre-blind testing informed the range of release rates that should be performed during blind testing solely to maximize the utility of the limited gas quantity available for this testing. One of our objectives was to establish the probability of detection curve, and it would not have been a prudent use of time or gas to perform releases at rates orders of magnitude below or above those needed to establish the PD curve. Additionally, the METEC facility also has engineering limitations with respect to the emission rate and duration. Testers at METEC are well informed of the range of releases the facility is capable of conducting and the challenges with maintaining higher emission rates for long durations.

- This test described "Equipment Group and Unit" localization, but without real equipment in the field, this is impossible to ascertain. The size of the "boxes" were not provided. More importantly, the lack of realistic atmospheric turbulence generated by real equipment was not testable.

  The abstract states that, "The test site was subdivided into 20 boxes (**109 × 83 m each**)."

  The lack of equipment-induced turbulence is a recognized limitation of this test site. We make no claims that we would expect the same results at a site with real equipment. In fact, we state in section 2.2:

  "Detection, quantification, and localization performance metrics for the GreenLITE™ system (or any continuous monitoring sensor) may be expected to vary in different environments due to differences in area, topography, meteorology, and on-ground infrastructure, and the authors acknowledge that results from similar testing performed at an operational oil and gas facility would likely differ from the results of this work."

  And in section 4.1:

"Additional testing to further assess the capabilities of GreenLITE™ for use in monitoring oil and gas facilities should include the presence of multiple emissions sources, allowed (known or expected) emission events, and higher emissions rates, as well as testing in a more relevant environment containing oil and gas equipment and infrastructure."

We have added text about turbulence in section 2.2.

- The detailed descriptions of other test protocols on pages 1 and 3 should be removed in favor of a simple reference to the Bell et al. 2023 test protocols. Those test protocols are the only element of METEC that are relevant for this study, since it was not performed at METEC or a similar site. Specifically, the Alden et al. 2019 study uses a different (not tomographic) approach.

  The Alden study was referenced not for its measurement approach but for comparison to its experimental design and the inherent limitations within: limited number of known potential release locations, longer release duration, known release times and durations, and almost complete lack of opportunity for false positive detections. That said, the difference in approach is irrelevant, as the CMP allows key performance parameters of disparate measurement technologies and approaches to be compared using a common set of metrics.

- The METEC protocols cited require time-to-detect metrics. These should be discussed in the paper. Similarly, the real-time automated reporting of detected events should be shown.

  Time-to-detect results are provided in Appendix E and referenced in section 3.1. We are unsure how to show real-time automated reporting in a manuscript.

- Page 5: What are the heights of the reflectors?

  From section 2.3:

  "Because of the flat site topography and lack of obstructions, all chord endpoints (transceiver optical heads and retroreflectors) were simply placed **two meters** above ground level, with optical heads mounted on custom pipe-frame structures and retroreflectors mounted on tripods."

- What are the heights of the controlled releases?

This information has been added to the introduction.

- In what ways were the controlled releases structured to mimic oil and gas equipment emissions?

  All releases were performed at a constant rate, as is done at METEC according to Bell at al. 2023. Similar to testing conducted at METEC under the CMP, the intent of the protocol is primarily to establish empirical detection limits of systems. As such, it is less important that the emission rates or equipment simulate emission rates from real operational equipment, but rather are selected to "map" the operational limits of the system under test. Since our focus was to establish the probability of detection curve, experiments were conducted in the range of release rates needed to develop this primary metric. If the primary goal of the study was to understand the absence/presence of bias in emission rate estimates from a system, then it would be more important that the experiments conducted included the full range of emission events expected to be found at real facilities, and to have sufficient data across many other independent parameters such as emission source height, location, composition, wind speed, wind variability, etc.

- What are the heights of the two greenlite instruments?

  From section 2.3:

  "Because of the flat site topography and lack of obstructions, all chord endpoints (transceiver optical heads and retroreflectors) were simply placed **two meters** above ground level, with optical heads mounted on custom pipe-frame structures and retroreflectors mounted on tripods."

- What calibration of these instruments is required prior to and during deployment?

  No pre-deployment calibration is required. The instruments periodically self-calibrate their laser wavelengths through the use of an internal gas cell with known gas composition and pressure. A zero-path calibration is also periodically performed to compensate for any electrical or optical frequency response changes over time, though analysis of data spanning several years shows this is not really necessary. We chose not to include this information in the manuscript because we do not expect it to be of much interest to most readers.

- What types of calibrations are needed (e.g., water vapor?)

Meteorological data from onsite sensors and publicly available sources are used in the retrieval to dry air mixing ratio. References containing details of the retrieval method, which accounts for water vapor and other atmospheric constituents, can be found in Pernini et al. 2002, which is cited in our manuscript.

- Page 7: the evaluation of the instrument as installed at the test site to the provide guidance on release rates to the testers invalidates the blind nature of this test, as described above.

  The CMP states:

  "One of the primary objectives of this protocol is to evaluate the Probability of Detection curve across a range of emission rates. Therefore, the Test Center will vary or extend emission rates in the test matrix during the testing to produce detection rates from near-zero to near-100% for the performers participating, and taking into account the range of environmental conditions tested."

  Limited non-blind testing was performed prior to blind testing to establish this range of emission rates, rather than doing so during blind testing as a means to conserve a limited supply of CH_4 gas. As has been repeatedly stated, release rates, locations, times, and durations were not known by the performer team for any of the blind releases.

  When conducting testing at METEC, the test facility gathers similar input from the participating testers, and unblinded experiments are often performed as a commissioning exercise when solutions are first installed.

- More information is needed on the flat-fielding correction. Was the flat-fielding correction conducted during a period of no releases? How often was this performed and under what circumstances?

  The flat-field correction was performed continuously, both during periods of no releases and during releases.

- Page 9: it appears this method precludes the possibility of more than one emission source on a monitored site. This should be discussed more fully in the assumptions noted in the blind testing sections above.

  While this is technically incorrect, it does highlight a limiting aspect of this testing in that it was know a priori that there could only be a single emission source. Unlike many other blind studies, where the locations of potential emissions sources were known a priori, the source in this study was placed at a random location within any of the 20 non-perimeter grid boxes shown in Figure 1. All boxes are assessed simultaneously for the presence or absence of a source. In theory, any of the 20 boxes could have a source, including multiple boxes at the same time. In reality, it was known that only one source was available at the time of testing. In section 4.1 we recommend future testing with multiple emission sources.

- Page 11: the presence of a high number of false positives for on-site and off-site sources should probably be shown in the abstract alongside the true positive results.

  The CMP states that emissions identified as originating from off-site are not to be classified as either true positive or false positive detections but are to be omitted from detection classification entirely. Our reported metrics conform to the protocol.

  The suggestion to include false positive detection results in the abstract has merit. However, we feel the context surrounding these results is important and cannot be adequately captured in the abstract. Therefore, we feel is it appropriate to leave the false positive results and related discussion in the body of the manuscript where the reader can get the full context.

- Page 13-15: the direct use of METEC terminology for equipment group and equipment unit is not merited here and is misleading. Without denoting the sizes of each of the “boxes” used, it is not justifiable to imply that equipment-unit localization was tested and proven. Furthermore, localization depends strongly on the turbulent interference imposed by physical oil and gas equipment being located on site. Therefore, it is not justifiable to make such a leap to oil and gas infrastructure.

  The box size is provided in the abstract. As previously stated, we used the CMP terminology for simplicity. For additional clarity, we have changed “Equipment Unit” to “Cell Unit” and “Equipment Group” to “Cell Group”.

- Pages 14 and 17: it was not made clear throughout the paper that area sources were being tested, and not point sources. This represents a broadly misleading feature of the article, which supposes to test point sources of oil and gas equipment. The diffuse vs. point source distinction needs to be made much more clear in the introduction, methods, and results sections. The results are very different between point and area sources (fig 7) and so this needs to be a main point of the paper's methods, results, and conclusions sections.

Both area and point sources were tested, as mentioned in section 3.3, the Figure 7 caption, and Appendix B. There was no intent to deceive through the use of diffuse sources, as there is value is evaluating our system's performance with regards to both release types. A small area diffuse source such as that used in our work is representative of an open-top liquid storage tank, a floating roof tank, or an underground pipe leak that diffuses and spreads through the soil before reaching the surface. The CMP does not specify that it may only be applied to point sources. We have added text to the introduction to make it clear that both diffuse and point releases were tested, but we disagree that this should be a main point of the paper.

- Page 21: off-site interference is a real issue in oil and gas monitoring. These events need to be classified as false positives regardless of origin.

We fully agree that off-site interference is a real issue in oil and gas monitoring, and that detection and attribution of these sources is critical. The fact that our system did exactly that is a credit to its effectiveness. However, the CMP states that emissions identified as originating from off-site are not to be classified as either true positive or false positive detections but are to be omitted from detection classification entirely. Our reported metrics conform to the protocol, and text has been added to clarify this.