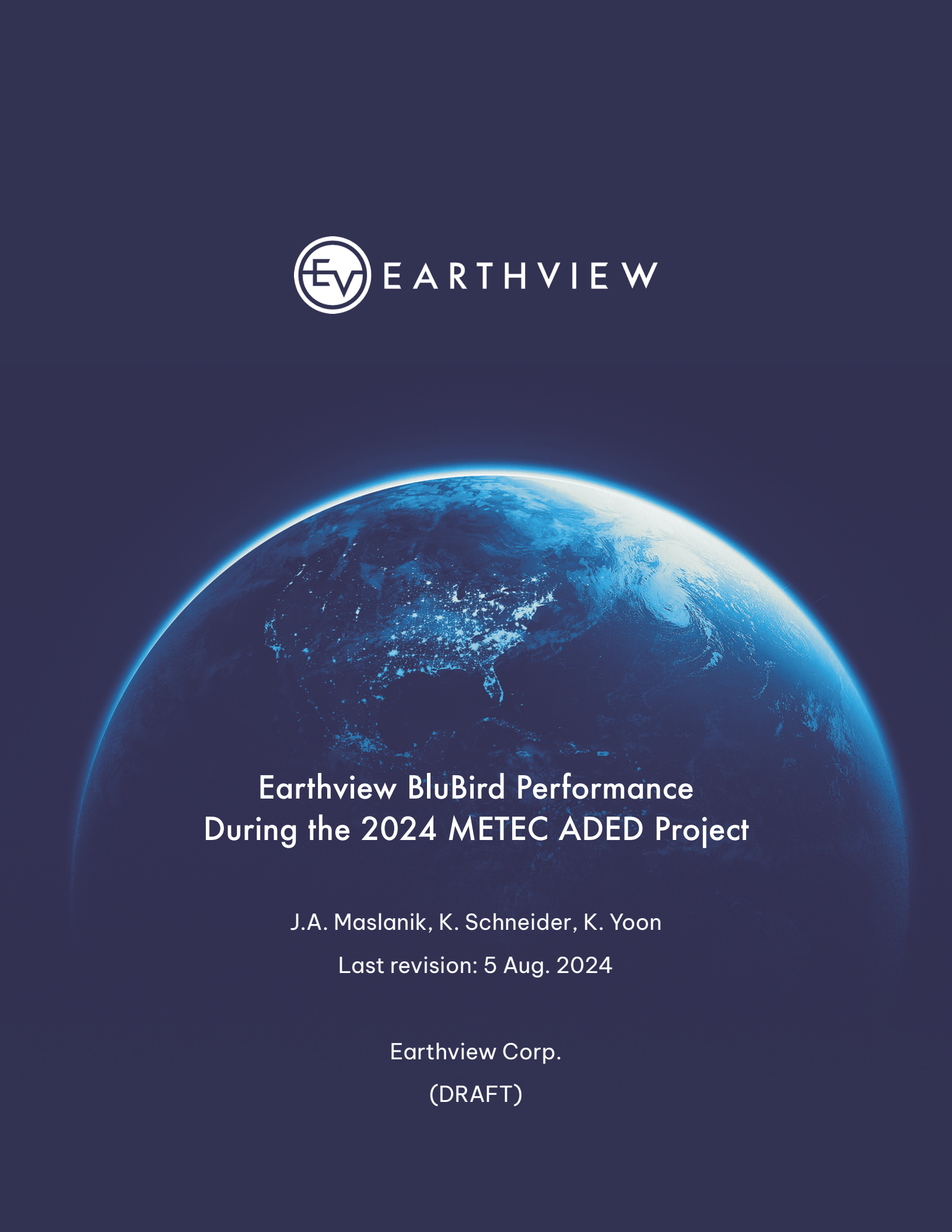

Earthview BluBird Performance During the 2024 METEC ADED Project

J.A. Maslanik, K. Schneider, K. Yoon

Last revision: 5 Aug. 2024

Earthview Corp.

(DRAFT)

Abstract: Earthview participated in the 2024 METEC single-blind ADED field testing during February-April 2024. The test period consisted of 775 individual releases of natural gas, grouped within 346 individual experiments. One hundred and five experiments consisted of a single release from the METEC test pad with the remaining 231 experiments consisting of multiple simultaneous releases from between 2 to 5 locations on the pad. Using METEC's standard grading criteria, Earthview detected 90% of the experiments and 69% of the total individual releases (536 of 775). Ninety-three percent of releases at 1000 g/hr or less were detected, as were 87% of releases of 400 g/hr or less. Leak event duration was accurately determined, with average durations of 229 minutes versus METEC's average of 224 minutes. For single-release events, BluBird correctly identified the specific equipment group (tanks, separators, wellheads, etc.) 49% of the time, and the general location on the pad 86% of the time. Sixty percent of the multi-release events were correctly identified as having at least 2 releases. Using an alternative grading criterion that assigns a true positive detection if a reported leak period overlaps with a METEC-defined leak period, the detection percentages increase to 97% for all experiments and for single-release experiments.

1. Background

This report summarizes preliminary results from Earthview's participation in field testing designed to assess the ability of emissions monitoring systems to detect and quantify methane leaks under a wide range of environmental conditions. Here, we focus on findings from an experiment carried out as part of the "Advancing Development of Emissions Detection" (ADED; https://energy.colostate.edu/metec/aded/) project by the Methane Emissions Technology Evaluation Center (METEC; https://energy.colostate.edu/metec/), operated by Colorado State University (Bell et al., 2023; Ilonze et al., 2024).

The ADED tests consisted of controlled releases of natural gas from 1 to 5 locations on the METEC field test facility (https://metec.colostate.edu/). METEC operators were

aware of the gas release timing and conditions but the participants were not (i.e., the tests were single-blind). The testing phase ran from Feb. 6th – April 29th 2024, with a total of 346 release periods, under a wide range of typical Colorado winter through spring conditions. Each of these 346 "experiments" consisted of from 1 to 5 simultaneous releases from any of 5 locations on the test pad. Eighty-seven percent of the events included more than one release source. Minimum and maximum release rates among all 776 releases as recorded by METEC were 81 g/hr and 6751 g/hr. The average methane release rate over all releases was 958 g/hr. Seventy-three percent of those releases were at rates less than 1000 g/hr, with 90% of the releases at less than 2.0 kg/hr. Release durations spanned 0.5 h to 8.2 h with an average duration of 3.2 hr.

As noted below, we also calculated the maximum release rate per experiment. These event-maximum release rates ranged from 128 g/hr to 6751 g/hr.

2. Implications of the standard METEC ADED participant assessment protocol

For the analysis presented here, we focus on the question, for how many of the 346 experiments did BluBird detect that there was a leak on the site? This is consistent with the main goal of the BluBird system, which is to determine whether there is an emission at a site, and to quantify and localize it in real time. The ADED protocol, however, requires that the technology identify individual detections for each release during multi-release experiments. In terms of assigning whether the leak was detected (a true positive [TP]) versus whether a leak was missed (a false negative [FN]), METEC's standard assessment methods do not differentiate between whether a monitoring technology entirely misses an experiment (i.e., no leak was detected at the site) versus whether the technology missed one release among five during an experiment. One FN would be assigned in either case.

This grading strategy can mask important differences among sensor technologies. Consider a set of four ADED experiments; one experiment has five simultaneous releases and the other three experiments have single releases.

Monitoring technology "A" successfully detects that a leak is present on the site during all four experiments but reports a single release for the five-release experiment. Monitoring Technology "B" detects all five of the simultaneous releases during the multi-release experiment but fails to detect the presence of any leaks during the other three single-release experiments. In other words, "A" caught all four of the experiments while "B" caught only one of the four experiments. "A" would be assigned four TP detections along with four FNs. Technology "B" would be graded as having five TPs and three FNs. If only the counts of TPs and FNs are reported, as is the case with METEC's ADED performance reports and publications to date, then one might conclude that Technology "B" performed better. However, it can be argued that, for at least some oil and gas operators, a monitoring technology such as Technology "A" that detects a higher proportion of individual leak events is more valuable than a technology that misses some individual events but can detect multiple, simultaneous releases. Again, this information is obscured in METEC's approach of simply reporting the totals of TPs and FNs.

With this in mind, for the results presented here, we adopt the approach used by Guerra et al. (2023) and consider each of the 346 experiments as a site-level leak event, with the maximum release rate for the experiment used for analysis.

2.1 Data

Earthview's emission detection procedure uses estimated emission rates calculated from measured methane concentrations. These emission rates are calculated automatically in real time from the array of BluBird sensor nodes at a site. The time series of emission rates is then analyzed using a moving time window to identify likely emission events. Detections and alerts are based on whether the estimated rates exceed a customer-defined threshold, with emission alerts issued within 30 minutes of the start of an event. Additionally, alerts can be triggered based on methane concentration alone, which provides an immediate notice of a potential problem on the site.

Analysis was carried out using version 3 of the preliminary set of ADED emission data provided to participants by METEC on 11 July 2024. This data set includes a variety of parameters such as METEC's release start and end times, metered release rates, and Earthview's reported detection information. METEC deleted several experiments from analysis due to overlaps with maintenance periods, resulting in 346 valid experiments.

Two subsets of the event data sets are analyzed here; both of which help address the confusion caused for performance assessment due to the presence of multiple releases during events. The first includes all the METEC events ("experiments") and assigns the maximum emission rate to the event if there were multiple releases during that event, as was done by Guerra et al. (2023). We refer to this data subset as the experiment-maximum rate data set. We also created a subset of the events that consists of only a single release (Single-Release Events). Results are also presented in terms of detection rates for all releases.

Results are presented mainly in terms of METEC's ADED protocol standard grading of true positives, false negatives and false positives. We also assess the performance when the classification of true positives and false negatives is based on whether a release period reported by METEC overlaps with a METEC release (see Section 3.1). This second approach is less conservative than the standard ADED protocol, resulting in a higher number of true positives.

3. Results

3.1. Event Detection

The capability of BluBird to detect the METEC releases is illustrated in Figure 1. These plots show typical sequences of METEC release periods (in black) and the release periods reported by BluBird (in blue). Each bar covers the period from reported leak start time to reported end time.

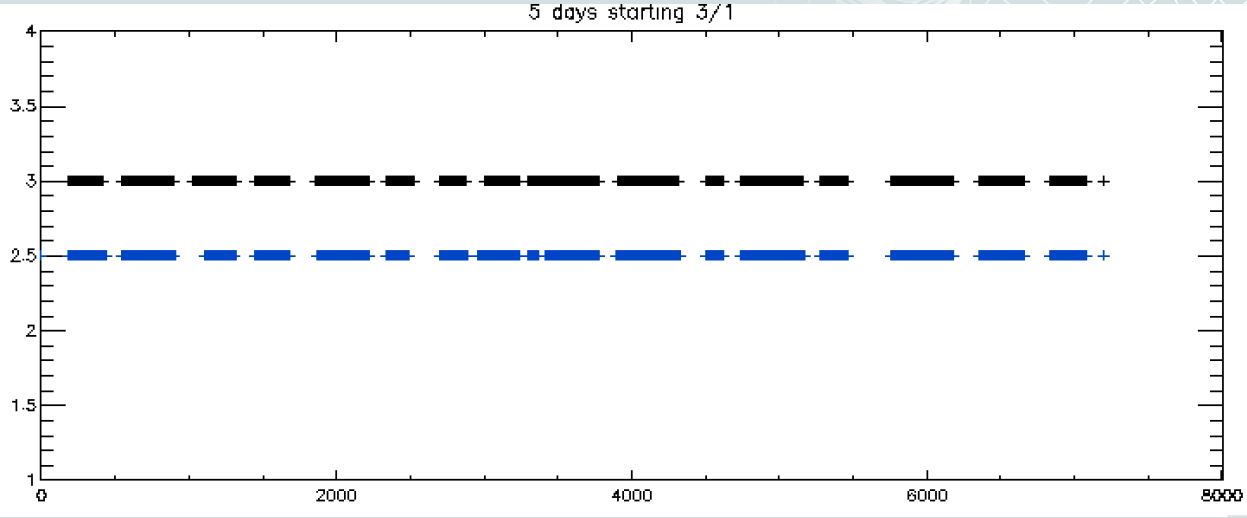

Figure 1a

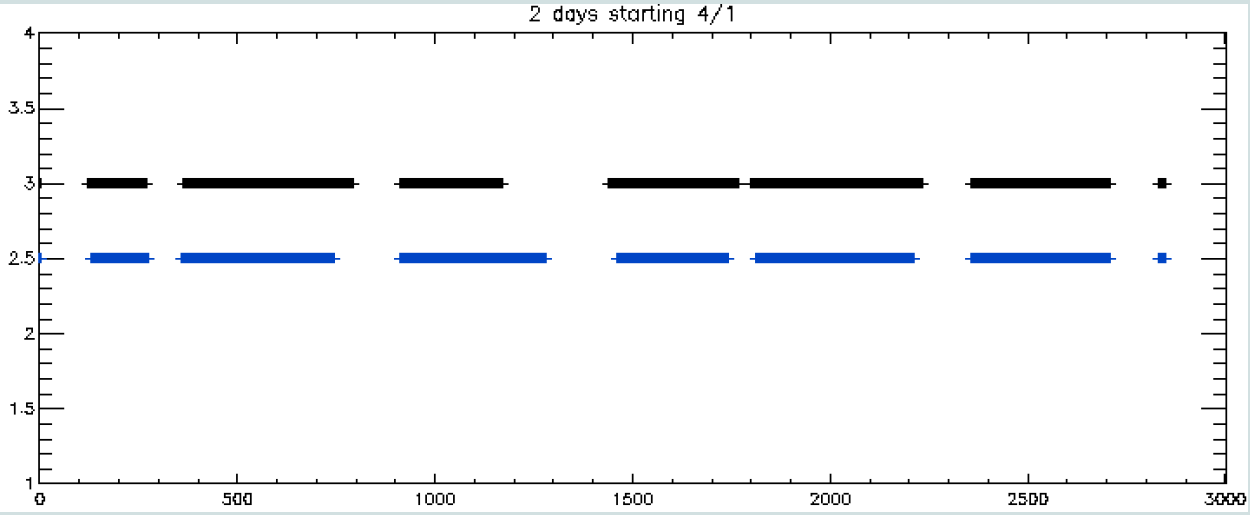

Figure 1b

Figure 1. Comparison of METEC release periods (in black) and corresponding BluBird-detected events (blue) spanning 5 days from 1 March (Figure 1a) and 3 days from 1 April. The x axis shows elapsed time in minutes. These examples are typical of results for the full study period.

Of the 346 events, there were 6 events where the BluBird release spanned more than one METEC release. These typically occurred when the METEC releases were spaced close together, such as the example in Figure 2 at around observations 0-200 and observations 1900-2100.

As noted earlier, we developed an alternative detection classification that is based on whether the Earthview-reported leak period (defined by the reported start and end time) overlaps with a single METEC-reported leak period. This classification approach essentially addresses the basic question of whether Earthview would have issued a leak alert to the customer in conjunction with a METEC release.

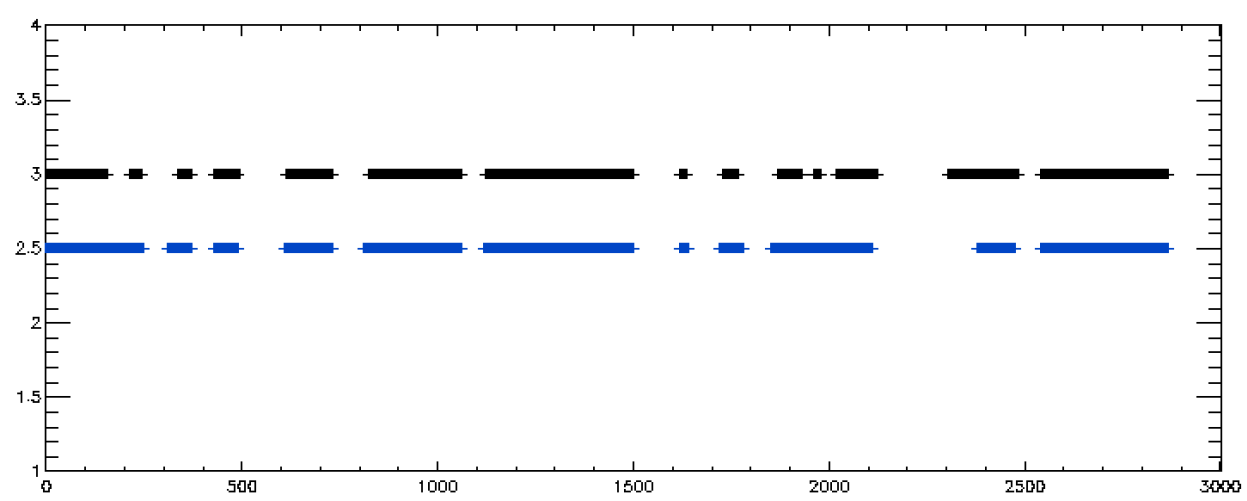

Figure 2. Example of a BluBird detection that overlaps multiple METEC releases.

3.1. Detection of All Releases

Earthview detected 69% of the 775 releases, according to the METEC grading criteria (i.e., 69% were assigned as True Positives).

3.2. Detection of Individual Experiments

BluBird detected 91% of individual experiments (314 of 346; Figure 3) based on METEC's standard grading. The detection rate was fairly consistent across all release rates (Figure 4). Ninety-three percent of experiments with maximum release rates of 1000 g/hr or less were detected, as were 87% of releases at 400 g/hr or less.

Using the "release overlap" criterion, the number of detections increases to 97% (335 of 346).

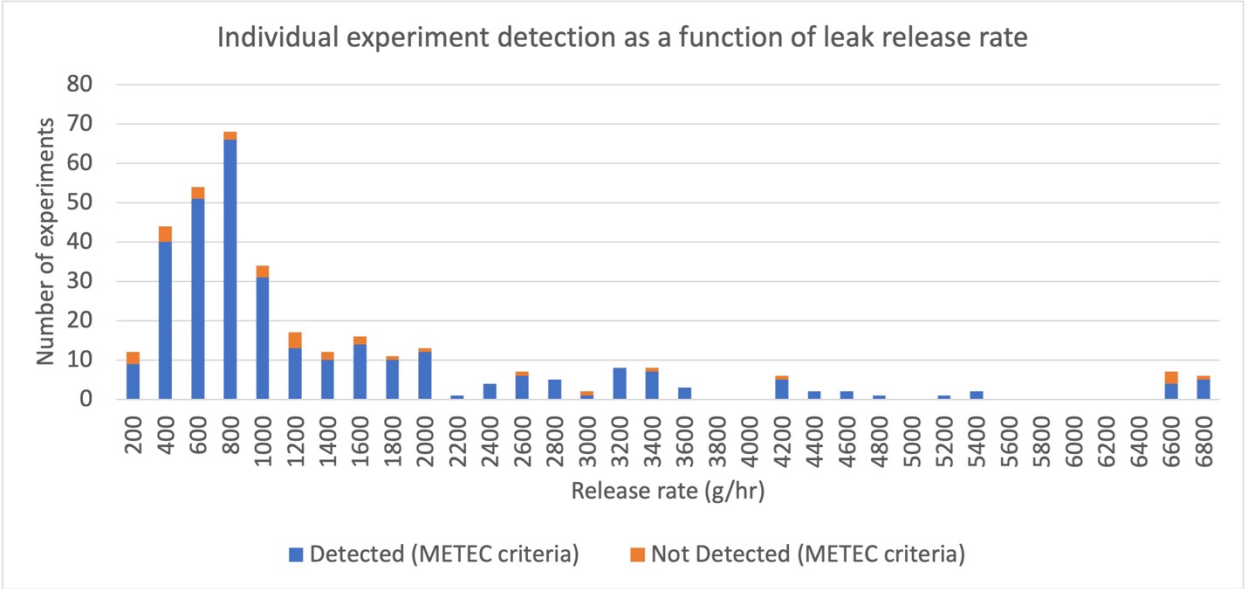

Figure 3. Number of detections per leak release rate, determined using METEC's standard criteria.

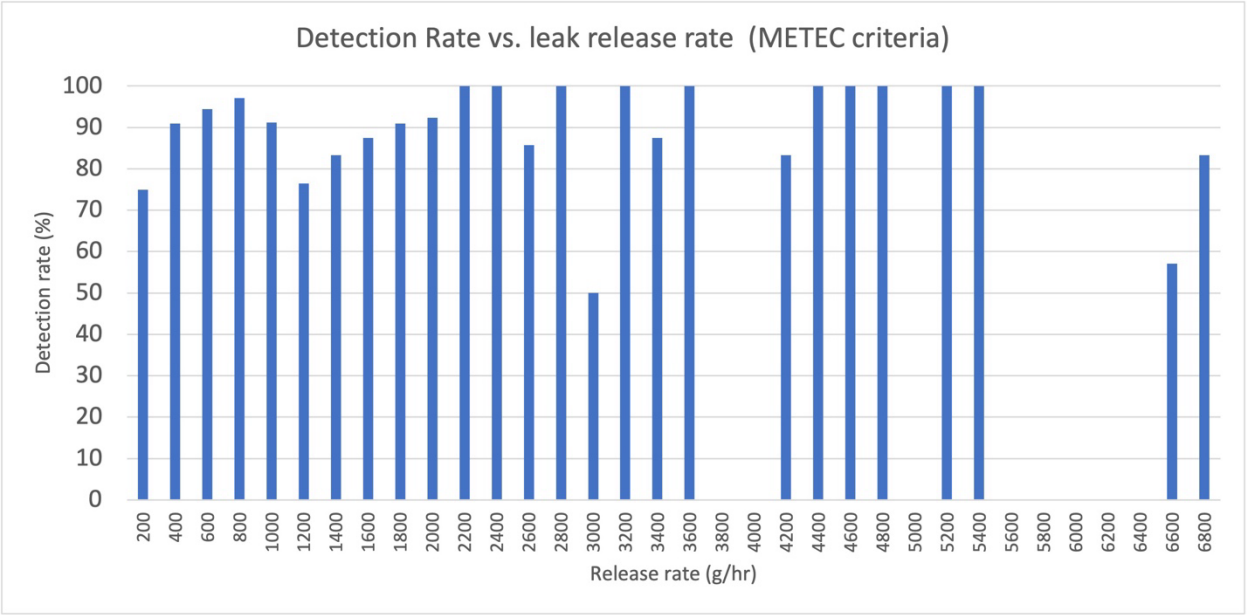

Figure 4. Percent detection per leak release rate, determined using METEC's standard criteria.

3.3. Detection rates for single-release experiments

It is worth considering whether the BluBird performance differed significantly depending on whether experiments consisted of multiple simultaneous releases or single releases. Using the standard METEC scoring, BluBird detected 85% (88 of 103) single-release experiments. This compares to the 91% detection rate for all multiple-release experiments.

Using the release overlap criterion, 100 of the 103 experiments (97%) were detected.

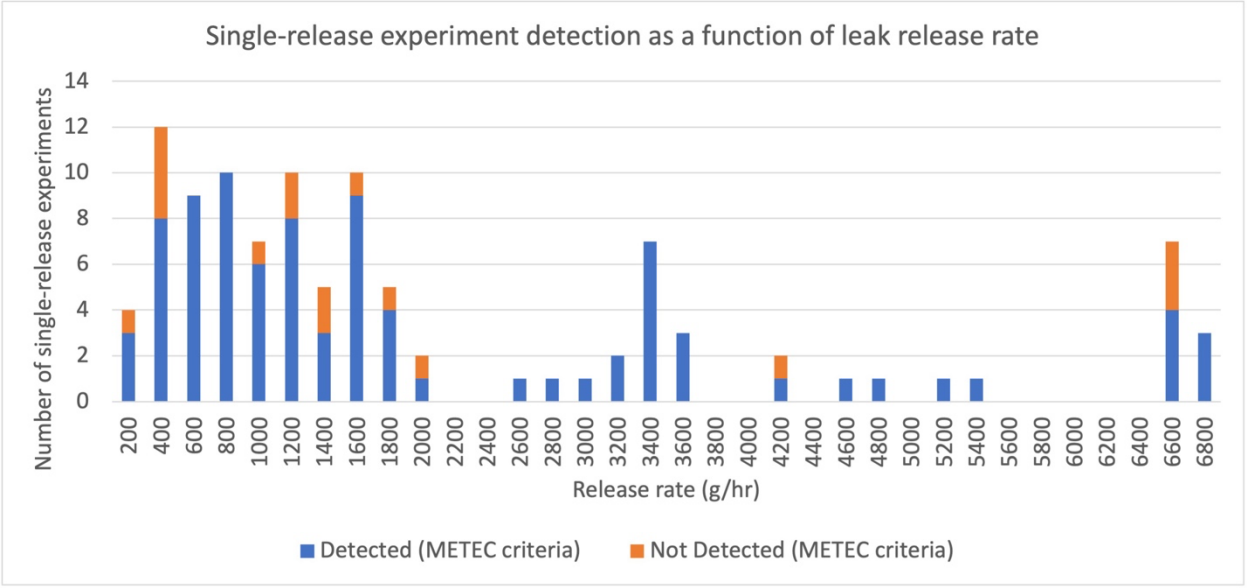

Figure 5. Number of detections per leak release rate for single-release experiments, determined using METEC's standard criteria.

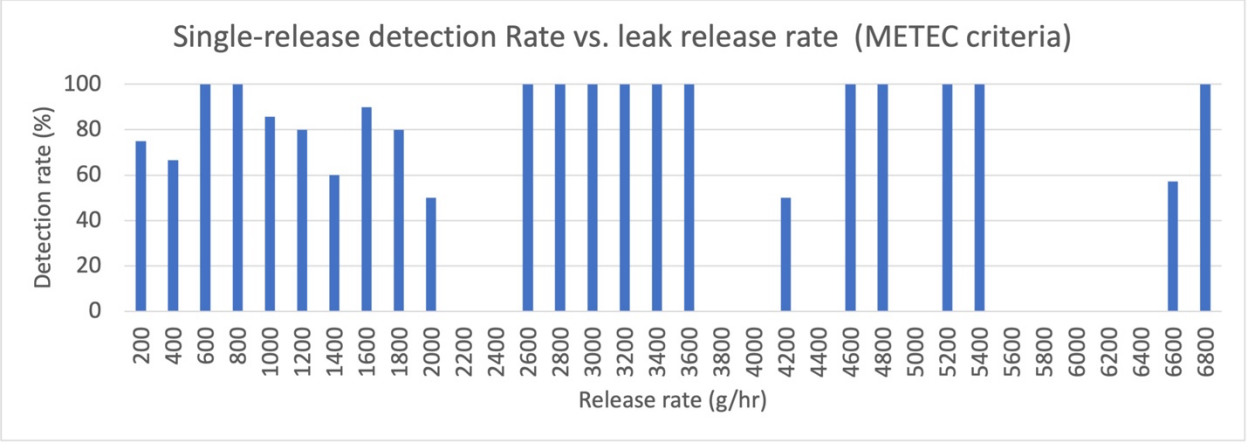

Figure 6. Percent detection per leak release rate for single-release experiments, determined using METEC's standard criteria.

3.4. Factors Affecting Detection

The detection rates do not appear to correlate well with emission rate. For example, in Figure 5, the lower detection percentage for the maximum release rates stand out, and seems counter-intuitive. Instead, detection success likely relates more closely related to leak duration and wind speed. Of the ten single-release experiments with release rates greater than 6000 g/hr, the three that were not detected had release durations of one hour or less (two of the three durations were less than one hour). Wind speed may also have been a factor, with 4 of the 15 missed releases corresponding to METEC-reported average wind speeds greater than 5 m/s. Twenty-eight percent of the missed experiments (9 of 32) had experiment durations of less than 1 hour, while 19% occurred with wind speeds greater than 5 m/s.

3.5. Limit of Detection (LoD) and Minimum Detection Level (MDL)

BluBird detected 3 of the 4 releases with rates between 100 and 200 g/hr, 4 of 4 releases for rates between 300 and 400 g/hr, 9 of 9 releases between 300 and 500 g/hr, and 29 of 30 releases between 300 and 1000 g/hr. We therefore conclude that for these test data, it is reasonable to claim a LoD of 200 g/hr and a MDL of at least 500 g/hr, with a conservative value of 1000 g/hr. This is consistent with Figure 7 (shown earlier), where 100% of the releases between 400 and 600 g/hr were detected.

3.6. Probability of Detection

As pointed out by MacGregor (2024), the standard definitions of Probability of Detection (PoD) are not necessarily particularly applicable to continuous monitoring systems. The typical application of logistic regression to define the 90% PoD of methane monitoring systems assumes that the main controlling factor is leak rate (e.g., Ilonze et al., 2024). However, for continuous monitoring devices that can detect low emission rates (around 1000 g/hr or lower), the more relevant factor is the probability that the methane plume from a leak will intersect a sensor. This intersection in turn is a function of wind direction, distance between sensor and source, placement of sensors, and number of sensors. For example, even a single BluBird sensor could achieve 90% PoD if leaks persist long enough so that the wind direction eventually aligns with the sources and sensor. We therefore agree with MacGregor that for continuous monitoring systems, PoD is essentially equivalent to the minimum

detection limit. However, in the following sections, to be consistent with other studies we consider PoD estimates based on release rates.

METEC defines PoD as "the fraction of binned test conditions (i.e., emission rate, release duration, etc.) classified as TP detections (i.e., TPs/TPs+FNs." (Ilonze et al., 2024). Therefore, using METEC's grading protocol, BluBird achieved a greater than 90% PoD for detection of individual events (314 of 346).

Using the logistic regression approach where "detection" is the dependent variable (0 vs. 1 for not detected and detected) and the predictor variable is experiment-maximum emission rate (e.g., Bell et al., 2023), the PoD reaches 90% at an emission rate of around 100 g/hr. In this case as well as for the other PoD calculations, the predictive power of the model is relatively weak, again suggesting that the detection rate is not a strong function of emission rates because the BluBird was able to detect leaks at nearly all the release rates used by METEC.

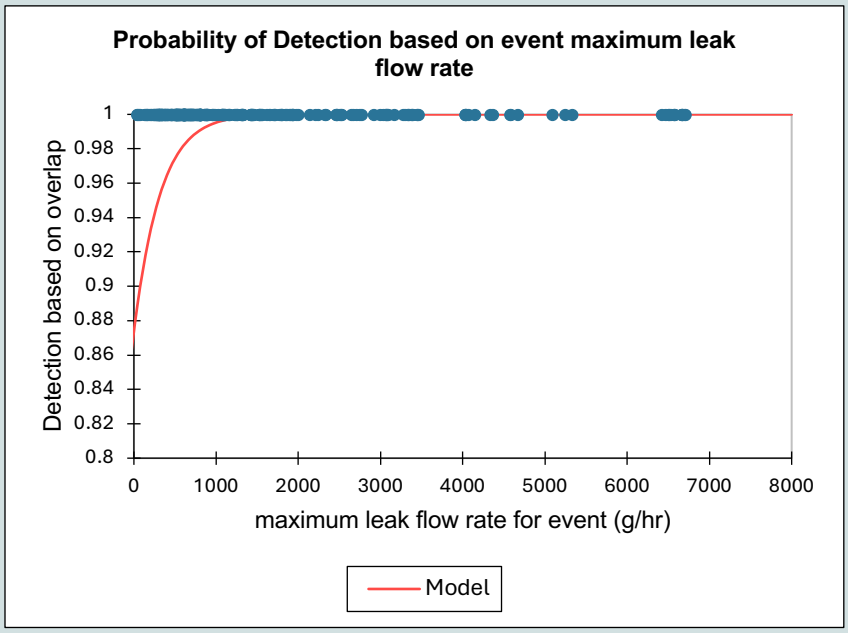

Figure 7. Probability of detection estimation using logistic regression applied to event maximum leak-flow rate (quantitative parameter) with overlap-based detection flag as the response parameter. The POD reaches 90% at a leak flow rate of 90 g/hr.

A third way of visualizing the probability of detection is to calculate the percent of time that release events were detected for individual ranges of release rates. These results were presented earlier in figures 4 and 6.

A review of the BluBird data for the missed experiments suggests two main causes for failure to detect releases. In a few cases, it appears likely that the methane plume was relatively narrow and aligned directly between two nodes. In these cases, it is very likely that the leaks would have been detected if the leak duration had been long enough so that a change in wind direction would have shifted the plume direction. In other cases, a review of the BluBird sensor-level data (resistances and derived methane concentrations) for the missed events reveals distinct signals typical of gas releases, but these were not positively classified by the automatic leak classification process.

3.7 Emission Source Location Identification

Using the single-release event data set, BluBird was scored by METEC as having correctly identified the source's equipment group or unit 76% of the time. BluBird was 86% accurate in providing information about the general location on the pad (whether the source was on the east or west side of the pad, for example).

3.8 False Positives

METEC's ADED standard grading protocol classified 164 of the BluBird detections as False Positives, of which only 10 arguably fit the practical definition of a false detection (i.e., a report of a leak when no leak is present). One hundred twenty of the 164 are noted in the ADED summary as being "Extra Reports" and 44 are assigned as "No Experiment Running." For the following reasons, it is reasonable to consider only the latter as being actual False Positives, in the sense of potentially sending an alert to a customer when there is actually no leak present. The Extra Report cases are cases where BluBird assigned an extra leak detection to multiple-leak events. For example, detections might have been reported for 3 individual releases when there might only have been 2 actual releases during that event. (This issue of multiple releases during ADED is discussed further in Section 4.) In these cases, grading such reports as False Positives is misleading in that BluBird actually detected the leak event that the extra report was associated with. In other words, this was not a false detection of an event

(and thus not a "false alarm"), but was a false detection of an additional release during an event that itself was correctly identified (i.e., assigned a True Positive).

Of the remaining 44 False Positives (the "No Experiment Running" cases), 10 appear to be actual False Positives (2 due to errors in time field entries and 8 due to misclassifications of events). The remaining assigned False Positives consist of 27 detections rejected due to early-than-allowed start times (earlier than 20.0 minutes but which overlap with a METEC release) and 7 detections that overlap more than one ADED release period. It is worth reiterating that each of these 27 in fact coincides in part with an actual ADED event.

4. Implications of ADED Performance Grading Protocol

The METEC ADED protocol requires that the technology report individual detections for each release during a multi-release experiment. For example, consider a set of 4 events (ADED "experiments"); 1 event has 5 simultaneous releases and the other 3 events have single releases. Monitoring technology "A" successfully detects that a leak is present on the site during all 4 events but reports a single release for the 5-release event. Monitoring Technology "B" detects all 5 releases during the multi-release event but fails to detect leaks during the other 3 single-release events. Technology "A" would be graded as having 4 True Positive detections but also 4 False Negatives. Technology "B" would be graded as having 5 True Positives and 3 False Negatives. If only the counts of True Positive and False Negatives are reported, then one might conclude that Technology "B" performed better, but it can be easily argued that a monitoring technology such as the BluBird system that detects a higher proportion of individual leak events is more valuable than a technology that misses some individual events but can detect multiple, simultaneous releases.

Second, we return to the issue that the ADED protocol excludes any detection report that has an event start time earlier than the official METEC start time by 20.0 minutes. This is meant to take into account some uncertainty in the exact timing of the natural gas releases. The BluBird automatic leak detection system uses a 10-minute moving time window to identify the start of leaks. If a leak is detected during that window,

the leak start time is assigned the earliest time during the window. Thus, the true start time could be anywhere within the 10-minute period, but the earliest start time will be what is recorded. This results in the BluBird start times being biased slightly early.

While this timing issue would seldom be of concern for normal field applications, it does have implications for METEC's ADED grading. Using the ADED 20.0-minute threshold results in 314 detections when applied to the event maximum-emission results (a 90% detection rate), but if the threshold is raised just slightly to 20.5 minutes, the assigned detection rate increases to 92%. So, if BluBird had used the center time of the 10-minute moving window for example, instead of the start time, the ADED-assigned detection rate would have improved. Other factors may have contributed to variations in determined start times, such as possible presence of residual methane on the site or recirculation of methane by wind patterns, particularly when the times between events were short. In such cases, the BluBird system would assume that such slightly elevated concentrations were the "leading edge" of the leak event, and thus have assigned an earlier start time.

For example, Figure 7 shows such a case, where a BluBird detections start earlier than METEC's reported start time (and also ends slightly later). Some of these events would be rejected by the standard ADED start-time check, and would be assigned as a False Negative or False Positive. This is despite the clear correspondence with the METEC releases. It is worth reiterating that BluBird would have alerted an operator that a leak was present for these events.

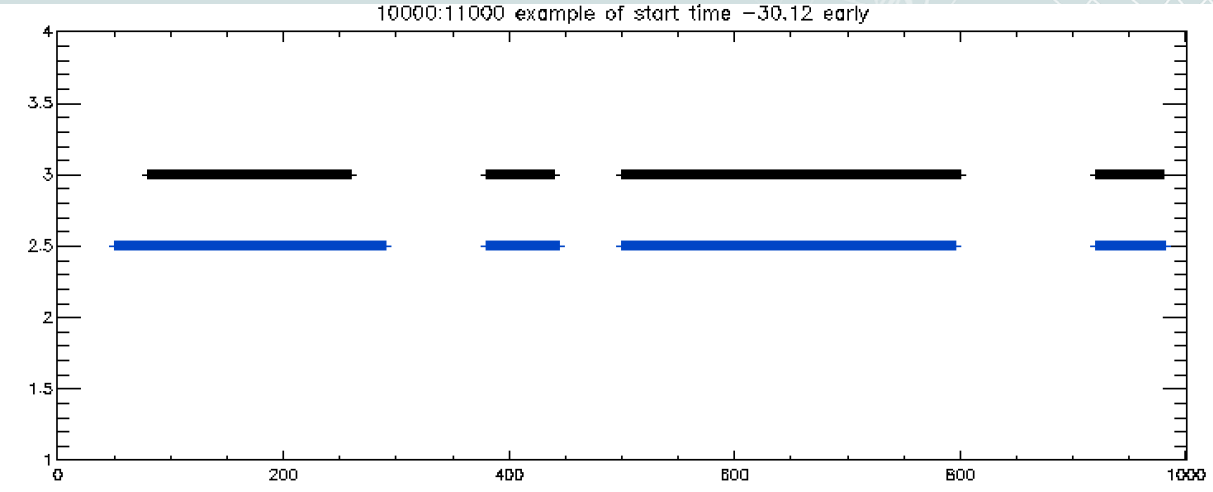

Figure 7. Examples of detections where the BluBird event starts noticeably earlier than the METEC-reported event, but which clearly corresponds to the METEC event.

The effect of this slight change in protocol on the number of detections and resulting probability of detection is significant. For example, using the maximum-release-rate individual experiment data, Earthview reported 335 of 346 leak periods that overlapped with actual METEC releases, or a detection rate of 97%.
For single-release experiments, 98% of the METEC release periods overlapped with a BluBird-reported leak period.

The effects of slightly relaxing the METEC protocol is also reflected in Figure 8. Changing METEC's allowable early start time from 20 minutes (Figure 8a) to 25 minutes (Figure 8b) results in a PoD greater than 90% for the full release-rate range.

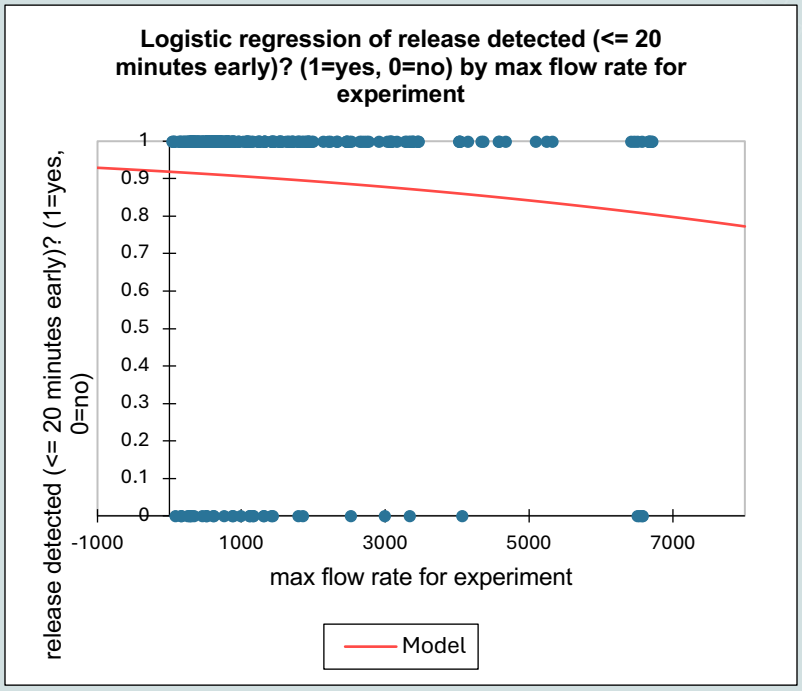

Figure 8a.

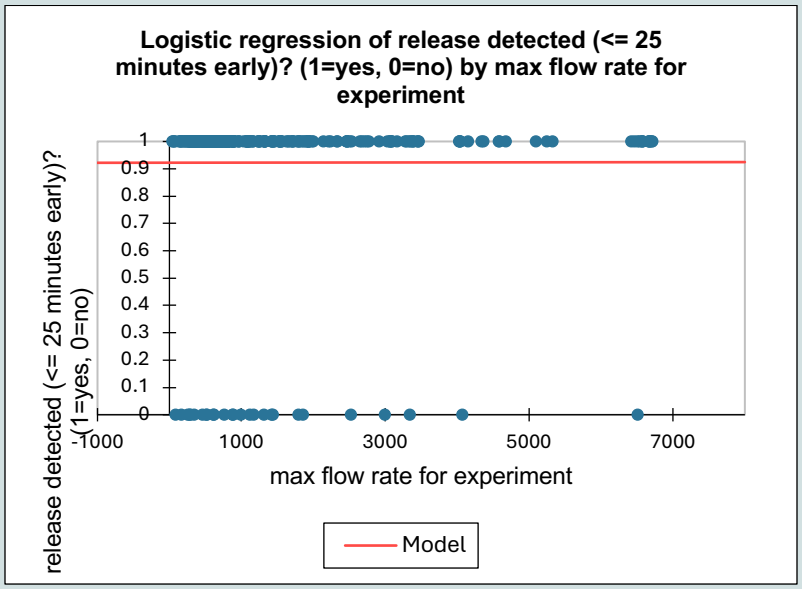

Figure 8b.

Figure 8. POD curves derived from the event maximum release rate data set, but with the METEC ADED protocol's start-time threshold applied (20 minutes for Figure 8a; 25 minutes for Figure 8b).

5. Comparison with ADED 2023 "Participant D" Results Using the Standard ADED Grading Protocol

It is valuable to consider how the BluBird performance stacks up against other detection technologies when METEC's standard grading protocol is applied. We can do this by making use of the ADED 2023 experiment results recently published by Ilonze et al. (2024) and comparing the BluBird results discussed here with those of the ADED 2023's apparent top performer - "Participant D" - available in Ilonze et al.'s supplementary materials.

For ADED 2023, there were 548 individual releases over 258 experiments (based on the Participant D results file) compared to 776 releases over 346 experiments for ADED 2024. Seventy percent of the experiments were multi-release experiments in both 2023 and 2024, although slightly more of the experiments consisted of 4 and 5 releases in 2024.

Participant D detected 73% of the individual releases during ADED 2023 (398 of 548) and 89.5% of the individual experiments (230 of 257). From our review of the 2023 data, this appears to be the best performance among the 9 participating technologies. As presented earlier, for ADED 2024 and using METEC's standard grading protocol, BluBird detected 69% of the individual releases (535 of 776) and 91% (314 of 346) of individual experiments. If the ADED protocol's early-start-time threshold was relaxed even slightly (from 20 minutes to 20.5 minutes), BluBird's percent detection of experiments would increase to 92%, with a comparable improvement in the detection rate for individual releases.

6. Emission Rate Quantification Accuracy

Considering the set of ADED experiments that consisted of single releases, we find that, for all single-release experiments detected, with 5 large outliers removed (BluBird-estimated rates >> actual release rates; leaving a total of 99 experiments), the average error and median error are -0.20 and -0.46 respectively (i.e., -20% and -46%). The errors ranged from -0.96 to 2.47. Eighty-five percent of the estimates are within +/- 100% of the actual rate, 58% are within 70%, 34% are within +/- 50%, and 15% are within +/- 20%. Seventy-five percent of the estimated rates are within one standard deviation of the reported rates. In terms of all releases (including multiple simultaneous releases), METEC's analysis of the 2024 ADED results yielded a mean quantification accuracy (absolute) of 98 g/hr and a mean accuracy (relative) of 0.72 times the actual flow rate.

A plot of BluBird-estimated release rates versus the actual METEC rates (Figure 9) reveals that the BluBird rates tend to overestimate the actual rates for lower rates, while underestimating the higher rates. The correlation is statistically significant at the 99% level, with a degree of scatter that is not too far off from that seen in the OTM-33A Picarro results from Edie et al. (2022) given above in Figure 10. We anticipate using these METEC data to further refine our response functions for converting MOS resistances to methane concentrations, and for improving our GPAQS emission rate estimation methodology.

Figure 9. Correlation of BluBird-measured emission rates versus known release rates for single-release experiments during METEC ADED 2024. Intercept is set to zero. Five outliers have been removed.

It is important to note that these results were obtained under actual field conditions, over a 3-month period, with a wide range of weather conditions, and at a site that is a reasonably good simulation of an actual production pad. Also, we made no attempt here to filter these data to select optimal wind conditions, such as was done by Benko et al. (2023) during testing of a similar continuous monitoring system.

Studies using inverse Gaussian plume modeling to calculate point emission rates from oil and gas facilities typically suggest accuracies of about +/- 0.3 to +/- 2.0, [(estimated emission rate-actual release rate]/[actual release rate]). Edie et al. (2020) reports an accuracy of +/- 0.7 for 95% of observations using the OTM 33A mobile approach carried out at the METEC facility, using releases from single and multiple sources. Sixty-eight percent of the rates are within 38% of the known release rates. Combining their test results with those of similar testing done by Robertson et al. (2017), Edie et al. found that greater than 85 % of the calculated emission rates are within ±50 % of the known rate, and 95 % of the estimated rates are within ±73 %. Figure x.3._1 (Figure 6 in Edie et al., 2020) illustrates the range of scatter seen in the combined data sets used (Christman data are those collected by Robertson et al. [2017]). Brantley et al. (2015) reported OTM-33A accuracy ranging from -60% to 52%, with 72% of estimated rates within +/- 30% of the actual rates. However, when not pre-screened, the accuracies ranged from -87% to 184%.

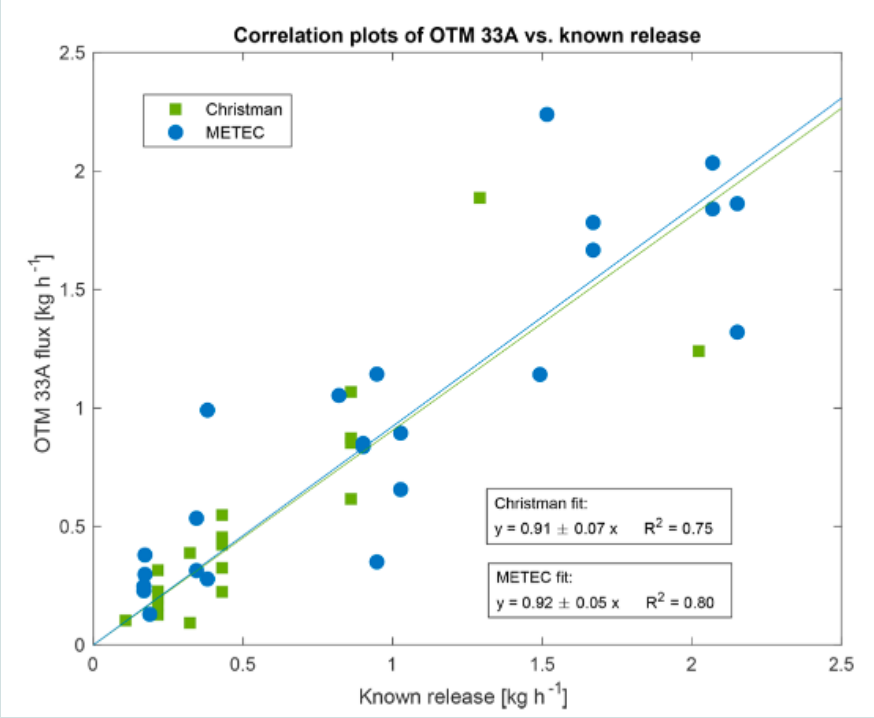

Figure 10. Correlation of OTM 33A-measured emission rates versus known release rates (From Edie et al. [2020], Figure 6).

EARTHVIEW

7. Summary

Results from the METEC ADED 2024 test period, using METEC's grading, show that the BluBird system detected 91% (314 of 346) of the ADED natural-gas release experiment, and 69% of all releases (536 of 775, including experiments with single and multiple releases). For individual experiments, using the maximum release rate per experiment, 93% of releases with rates of 1000 g/hr or less were detected, as were 87% of releases of 400 g/hr or less

When considering only the experiments with one gas release, 85% of these single-release experiments were detected, with 95% detection for rates between 500 and 1000 g/hr. The estimated Level of Detection and Minimum Detection Level for these experiments is 200 g/hr and less than 1000 g/hr, respectively. Source locations were correctly identified 76% of the time, with more general location on the pad identified 86% of the time.

Using an alternative grading criterion that assigns a true positive detection if a reported leak overlaps with an actual leak, the detection percentages increase to 97% for all experiments and for single-release experiments.

Probability of Detection (PoD), defined as the fraction of True Positives divided by number of experiments, was 91%. A 90% PoD, based on fitting a logistic regression to detection vs. release rate, was achieved at a rate of approximately 100 g/hr. Given the sensitivity of the BluBird instrument, for the release rates used during the ADED experiment the PoD is more likely a function of other factors such as gas release duration and wind speed rather than release rate.

If METEC's performance grading criteria is relaxed somewhat to consider whether a reported detection overlaps with a METEC release (with the main effect of allowing earlier start times for detections), then the BluBird detection rate increases further. With this criterion, the detection rate for all experiments increases to 97%, and 98% for single-release experiments.

BluBird's performance for ADED 2024 compares well with that of the apparent top performer in ADED 2023, based on data recently published by METEC and using the standard METEC grading protocol. The percent detection of experiments was slightly higher for BluBird, while detection of individual releases was slightly lower.

Estimated emission rates for the single-release experiments yielded an average error and median error of -20% and -46% (i.e., BluBird tended to underestimate the true release rates). These errors are consistent with the general performance seen in other similar applications of inverse Gaussian plume modeling for emission rate estimation from point leak sources on oil and gas pads.

Some aspects of the METEC ADED protocol for evaluating test participant performance were discussed, including implications of multiple releases during individual experiments. We recommend that METEC consider adding to their standard analysis an additional step that considers detection of individual experiments. Another option would be to classify false negatives into two categories; a false negative that meant that an experiment was entirely missed (an "Experiment False Negative", say), and a false negative that meant that one of the multiple releases during an experiment was missed (a "Release False Negative"). Performance metrics would then be calculated separately for these two categories. In terms of leak detection and repair, an Experiment False Negative is likely to be much more significant than a Release False Negative.

EARTHVIEW

References

Bell, C., C. Ilonze, A. Duggan, and D. Zimmerle, 2023. Performance of continuous emission monitoring solutions under a single-blind controlled testing protocol. Environ. Sci. and Technol., 57, 5794-5805.

Edie, R., A.M. Robertson, R.A. Field, J. Soltis, D.A. Snare, D. Zimmerle, C.S. Bell, T.L. Vaughn and S.M. Murphy, 2020. Constraining the accuracy of flux estimates using OTM 33A. Atmos. Meas. Tech., 13, 341-353, https://doi.org/10.5194/amt-13-341-2020.

Guerra, J.E., N. Eichenlaub, and D. Solomon, 2023. Event-based analysis of performance of Project Canary's continuous monitoring systems.

Project Canary, https://www.projectcanary.com/blog/event-based-analysis-of-performance-by-project-canarys-continuous-monitoring-systems/

Ilonze, C., E. Emerson, A. Duggan, and D. Zimmerle, 2024. Assessing the progress of the performance of continuous monitoring solutions under a single-blind controlled test protocol. Environ. Sci. Technol. Publ. date June 12, 2024, http://doi.org/10.1021/acs.est.3c08511, https://pubs.acs.org/doi/10.1021/acs.est.3c08511.

MacGregor, A., 2024. Why Probability of Detection (PoD) is a flawed metric for evaluating continuous monitoring systems. https://www.qubeiot.com/resources-news/why-probability-of-detection-is-a-flawed-metric-for-evaluating-continuous-monitoring-systems