Validation of the Fugitive Emission Distributed Sampling (FEDS) system: A mobile, multi-inlet system for continuous emissions monitoring

Jacob T. Shaw[1], Neil Howes[1], Jessica Connolly[1], Dragos E. Buculei[1], Jamie Ryan[1a], Jon Helmore[1], Nigel
5 Yarrow[1], David Butterfield[1], Fabrizio Innocenti[1], Rod Robinson[1]

[1]National Physical Laboratory, Hampton Road, Teddington, Middlesex, TW11 0LW, UK.
[a]Present address: Met Office, Fitzroy Road, Exeter, EX1 3PB, UK.

Correspondence to: Rod Robinson (rod.robinson@npl.co.uk)

Abstract. The National Physical Laboratory (NPL) has developed and trialled a mobile and remotely-operated Fugitive
10 Emission Distributed Sampling (FEDS) system for continuous measurements of emissions at the facility spatial scale. FEDS
is capable of both locating and quantifying emission sources over long-term periods and has been deployed at sites around the
UK to monitor methane emissions from the natural gas network, landfill, and waste treatment. This work presents validation
activities using a controlled release facility (CRF) to test the performance of the measurement system and two reverse
dispersion models (Airviro and WindTrax) for emission quantification. Emissions were quantified using prior knowledge of
15 release timings as well as in the absence of this knowledge. High variability in wind direction was shown to negatively impact
emission quantification accuracy (especially for Airviro). Emission results were improved by removing periods of high wind
variability (low wind persistence) from analysis. Both models performed better when using daily-averaging periods for
emissions (Airviro RMSE = 0.37 kg h^{-1}; WindTrax = 0.29 kg h^{-1}) over shorter averaging periods, such as hourly data (Airviro
RMSE = 0.77 kg h^{-1}; WindTrax = 2.19 kg h^{-1}). Emission rates were shown to be sensitive to the specified source release height
20 for both models, with discrepancies in model release height relative to the true release height of more than 0.5 m yielding less
accurate results. Furthermore, it was shown that emission results were less accurate when using concentration data input from
fewer sampling locations, although it would be remiss to recommend a minimum density of sampling locations for a given
area based on a single validation study.

Keywords: Methane, emission, network, dispersion, validation, continuous, long-term

1 Introduction

There is growing awareness of the need to reduce global anthropogenic emissions of methane (CH_4), a powerful greenhouse gas, to meet the goals of the Paris Agreement (Nisbet et al., 2020; 2022a; Ocko et al., 2021; Tollerson, 2022). The Global Methane Pledge, signed by over 100 nations at the 26th United Nations Climate Change Conference of the Parties (COP26) in 2021, requires signatories to cut their CH_4 emissions by 30% by 2030 (European Commission, 2021). Yet only 13% of global CH_4 emissions are currently included within direct mitigation policies (Olczak et al., 2023). CH_4 mitigation requires efficient and transparent monitoring of CH_4 emissions, both to ensure effective reductions are taking place and to ensure public confidence in strategies (Erland et al., 2022; Connor et al., 2024). However, monitoring CH_4 emissions is extremely challenging, especially across the diverse spatial, temporal, and emission magnitude scales required to guarantee accuracy and reliability, and to fully capture all known and unknown emission sources (Cusworth et al., 2020; Nisbet, 2022b; Wang et al., 2022).

Accurate monitoring of facility-scale CH_4 emissions typically requires measurements of atmospheric CH_4 around the facility of interest. There are many approaches to sampling atmospheric CH_4, whether that be from single-point stationary sensors, from a network of sensors, or from sensors mounted on a plethora of mobile platform options, including people, bicycles, ground-based vehicles, unmanned aerial vehicles (UAVs), manned aircraft, or satellites (e.g., Fox et al., 2019; Shaw et al., 2021; Erland et al., 2022; Jacob et al., 2022; Allen, 2023; Yang et al., 2023; Connor et al., 2024). Mobile platforms offer advantages in spatial coverage and can be used to target individual emissions sources for direct sampling. However, using mobile platforms can be labour intensive and therefore long-term or continuous deployment can be difficult. Single-point stationary sampling may have limited spatial coverage, but this can be alleviated by using a network of stationary sensors. However, installing networks for continuous monitoring can be prohibitive due to the expense of high precision analysers and a lack of reliable and high-performance low-cost alternatives (Fox et al., 2019).

Connor et al. (2024) presented a framework for monitoring CH_4 emissions highlighting the complexity of the monitoring landscape. Connor et al. (2024) emphasised that monitoring methods should adhere to basic principles of metrology and quality assurance in order to meet their data reporting requirements. Where possible, standardised methods should be used for formal reporting of CH_4 emissions but validation activities in controlled scenarios are an important aspect of achieving trust and confidence in any developing (or even standardised) method.

Kumar et al. (2022) presented results from several controlled-release validation experiments on a CH_4 monitoring network consisting of seven high-precision gas analysers (of various types) connected to 16 sampling inlets. Using multiple spectroscopic analysers allowed for the potential to monitor both the emission plume and the CH_4 background simultaneously (regardless of wind direction). However, using multiple expensive analysers would typically be beyond the scope of routine emission monitoring especially in the context of the many facilities in operation nationally and globally. Riddick et al. (2022) used a network of four calibrated low-cost sensors positioned 30 m from emissions of known rates over two days. The low-cost sensors were able to detect CH_4 emissions of 84 g h^{-1} for only 40% of the time. Emissions estimated using reverse

dispersion models were found to be only within ±50% of the release rate on a 1-day average. In addition, there are concerns that low-cost CH_4 sensors may be unreliable due to cross interference with parameters other than CH_4 (Riddick et al., 2022;
60   2020) and may only be applicable in limited environmental conditions (Rivera-Martinez et al., 2024).

An alternative solution to using multiple expensive spectroscopic analysers or a network of (currently) imprecise and unreliable low-cost sensors is to use a single expensive analyser connected to multiple sampling locations. The use of multiple sampling locations gives a few direct advantages over single-point monitoring. Firstly, multiple points-of-interest may be monitored in quick succession; separate emission sources may therefore be disaggregated from greater spatial data. Secondly,
65   multi-point monitoring is less dependent on wind direction (and wind speed) whereas a single-point monitoring system would be reliant on a potentially narrow range of wind directions across which emissions can be captured; wind directions outside of this range might result in measuring air masses which have not travelled over the source of interest (Shaw et al., 2020). Single-point monitoring is therefore much more likely to miss transient and irregular emission events from non-continuous sources due to imperfect wind conditions. With all fence-line monitoring approaches, and even with multiple sampling locations, there
70   still remains the possibility of missing emissions. Emissions may be missed, for example, if gaseous emissions are transported above the height of the sampling inlets, which can occur if the emission source emits gases at a higher temperature than the ambient atmosphere. This may be mitigated, of course, by placing inlets at multiple heights above the ground.

In this work, we evaluate the performance of the Fugitive Emission Distributed Sampling (FEDS) system, developed at NPL, for detecting, locating, and quantifying CH_4 emission sources. Multiple known and controlled releases of CH_4 were
75   used to evaluate and validate the FEDS performance. The FEDS method is intended for the continuous but cost-effective monitoring of CH_4 emissions. Furthermore, FEDS is envisioned to be accessible to non-academic users and this is exemplified through the use of commercially-available software for the direct quantification of CH_4 emissions. We also present results from a custom-built Gaussian plume model which, whilst less likely to be accessible to commercial users, provide additional context for validation of emission quantification performance.

80

2 Methodology

The FEDS method uses a mobile weatherproof station housing a high-performance gas analyser for the measurement of atmospheric trace gas concentrations. In principle, any high-performance analyser with sufficient capability for high resolution measurement of trace gases may be used. For the purposes of the work described here, an ABB Ltd. Los Gatos Research Inc.

85   (LGR) Fast Methane-Ethane Analyzer (FMEA) was used to measure atmospheric methane (CH_4) and ethane mole fraction data. A multiple-inlet unit (MIU; or multiplexer) is used to switch between up to 15 sampling inlets, located strategically around a facility to both optimise proximity to known emission sources and to capture emissions across a full range of possible wind directions (i.e., covering 360° around the target facility). Each inlet location is sampled in a pre-determined sequence providing discrete atmospheric trace gas measurements at high precision (i.e., nmol mol^{-1} (SI-preferred), or parts-per-billion

90   (ppb; used throughout this work) for CH_4) across a wide spatial area at least once every hour. In the current setup, a sixteenth sampling inlet is reserved for a single-point reference gas standard containing a known concentration of CH_4. Sampling inlets were protected from the weather via solar radiation shields attached at heights of ~2 m above the ground. Synflex 1300 tubing (polyethylene/aluminium composite) was used to connect sampling inlets to an external box on the FEDS trailer. Internally (within the trailer) all tubing was polytetrafluoroethylene (PTFE). Groups of four sample lines were connected to one of four

95   vacuum pumps, each with a specified throughput of 20 L min^{-1}. This created a draw of up to 5 L min^{-1} (realistically 2-3 L min^{-1}) at each sampling inlet location, depending on the relative distances of the four tubing lengths. The MIU was connected to each sampling inlet individually, as well as to the reference gas standard.

The central station also contains a 10 m telescopic hydraulic mast for the measurement of local meteorology (wind speed, wind direction, and air temperature). Meteorological data is also collected from the trailer roof to acquire air temperature

100   data at a second height. The acquisition of trace gas and meteorological data can be controlled and monitored remotely. Meteorology data was recorded by a CR6 datalogger (Campbell Scientific). Both the LGR FMEA and the CR6 datalogger were networked via a 3G cellular router for remote access. The 3G network in the UK was sufficient for cellular reception for the duration of the measurements in this work (although the router has since been upgraded for use with the 4G network).

Figure 1 shows a schematic of the FEDS facility used in this work. Measurements for the work presented here took

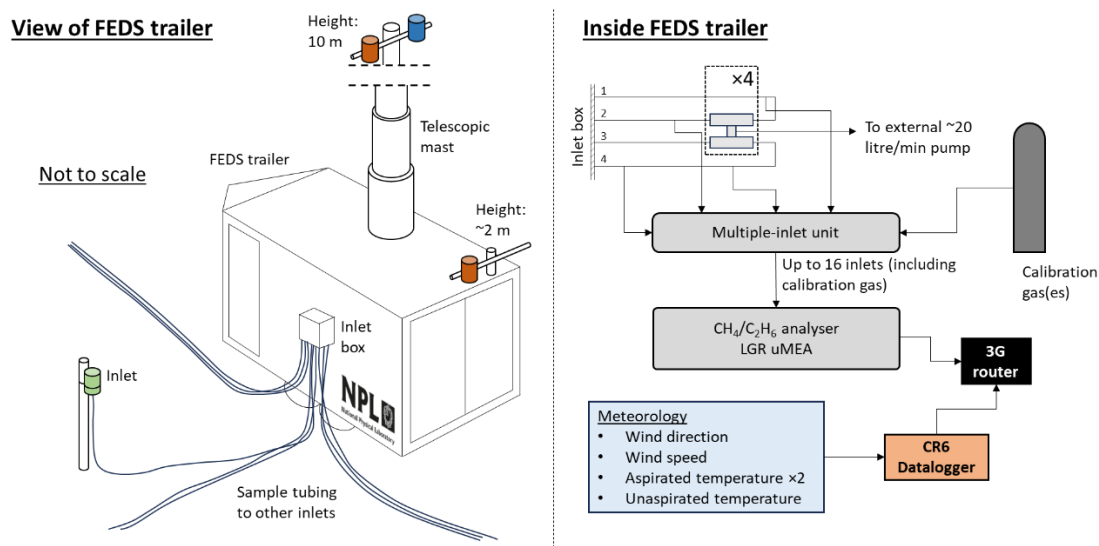105   place between 21st March and 8th April 2022.

Figure 1. Schematic showing the FEDS facility. LGR FMEA = Los Gatos Research Inc. Fast Methane-Ethane Analyzer. The heights on the FEDS trailer are defined as heights above the ground.

2.1 Methane mole fraction measurements

For the work presented here, an LGR FMEA was used to measure mole fractions of atmospheric CH_4 between 21st March and 8th April 2022. Atmospheric ethane (C_2H_6) mole fraction measurements were also recorded by the FMEA but C_2H_6 data were not assessed in this work. The FMEA has a reported precision of 2 ppb CH_4 (1σ) at 1 s and 0.6 ppb CH_4 (1σ) at 10 s, over a measurement range of 0 – 10,000 parts-per-million (ppm) CH_4 (Table 1). Whilst the FMEA normally operated at 1 Hz, a software issue led to a slow and steady degradation of sampling frequency over time (tests showed that measurements were made at roughly 0.25 Hz after approximately two weeks, and at roughly 0.15 Hz after approximately five weeks). Periodic cycling of the instrument power temporarily resolved the software issue and allowed the analyser to return to a sampling frequency of 1 Hz albeit with the same degradation pattern. The software issue was not expected to have impacted the accuracy of measured CH_4, although the reduced number of datapoints taken at each sampling location increased the associated standard error.

The MIU switched between sampling lines every 4 minutes so that CH_4 measurements were taken from each of the 15 inlet locations at least once every hour. Even whilst inactive on the MIU, the pumps were used to flush air through the sampling lines to maintain constant flow. However, the first period of data following switching to a new sampling line were considered invalid to remove the possibility of the FMEA measuring air drawn from the previous sampling line (e.g., due to non-instantaneous cell flushing time). The sensitivity of quantified emissions to this time period (referred to as the sampling lag time) was tested to determine the optimal time period to invalidate or exclude data (see Appendix C). Mole fraction data from each sampling location were averaged over hourly intervals to provide one single value per sampling location per hour to the emission quantification model(s).

130   A single reference gas standard (BOC Ltd.) with a nominal CH_4 mole fraction of 50.0 ($\pm$ 0.5) ppm was sampled by the gas analyser periodically to ensure traceability of measurements. The gas standard was sampled once per day on each of 24th, 28th, 29th and 31st March and 1st and 5th April, and twice on each of 30th March and 6th April, for four minutes each time. The first 60 seconds of data were removed (as above when switching over sampling inlets) to preclude the possibility of measuring air in the cell sampled from elsewhere (i.e., one of the other sampling inlet locations). The mean CH_4 mole fraction measured when sampling from the gas standard was 50.22 ($\pm$ 0.04) ppm across all samples, with a median value of 50.23 ppm

135 and a range 50.11 to 50.29 ppm. Analyser settings were not adjusted following sampling of the gas standard as all of the measured mole fractions were well within the uncertainty range of the reference gas standard. In practice, it would be preferential to employ a calibration scheme with more than one single standard reference point to account for any non-linearity in analyser response across a range of CH_4 mole fractions. However, the consistency between the analyser response and the calibration standard provides enough confidence that the data was of high precision for the duration of this validation campaign.

140   CH_4 mixing ratios measured whilst the flow into the optical cell was less than zero (which occurred during initial instrument start up and infrequently as sampling inlets were switched) and when the cell pressure was below 298.5 Torr (the target cell pressure) were removed before analysis due to inaccurate readings (i.e., mixing ratios well below the atmospheric CH_4 background).

2.2 Meteorology

145 Wind speed and wind direction were measured using a WindSonic anemometer (Gill Instruments) deployed at 10 m above ground on the hydraulic mast. Manufacturer-reported specifications are detailed in Table 1. Atmospheric temperature was measured at two heights using PT100 platinum temperature probes (R. M. Young Company). Temperature was measured at 2 m above ground in both an aspirated and unaspirated radiation shield and at 10 m above ground in an aspirated radiation shield only. The PT100 temperature probes were calibrated against a National Institute of Standards and Technology (NIST) traceable

150 thermometer by R. M. Young Company on 6th August 2020. All meteorological data were recorded as one-minute average data. Meteorology data were averaged across hourly periods for input into the emission quantification model(s).

Table 1. Manufacturer-reported specification for measured parameters.

Parameter	Accuracy	Resolution	Response time
Wind speed	±2% @ 12 m s^{-1}	0.01 m s^{-1}	0.25 seconds
Wind direction	±3° @ 12 m s^{-1} [a]	1°	0.25 seconds
Temperature	±0.3 °C @ 0 °C ±0.1 °C with NIST calibration		
Methane (CH$_4$)	2 ppb (1σ) @ 1 s 0.6 ppb (1σ) @ 10 s	0.01 ppb	1 Hz [b]

155   [a] Additional uncertainty in wind data is anticipated as the anemometer was only roughly aligned to the north (visually using a compass) and hence the accuracy of wind direction data quoted here should be treated as a minimum value only.

[b] A software issue on the FMEA led to the steady degradation of sampling frequency of CH$_4$ mole fractions over time. Periodic recycling of instrument power resolved this issue, but reduced data collection did occur.

160   The stability of local wind conditions can affect the accurate measurement of emissions and derivation of emission rates, particularly when averaging concentration measurements over periods of time during which the wind conditions were variable. This is partially because the average wind direction over a period of time may not be representative of the true range in wind directions during that period. For example, 30 minutes of wind from the east (90°) and 30 minutes of wind from the south (180°) would yield an hourly-average wind direction from the south southeast (135°). In reality, the wind did not blow from

165   the southeast, and enhancements in methane concentration may only have been measured at sampling locations directly to the west (when the wind was from the east) and to the north (when the wind was from the south) of the emission source, and not at locations to the northwest. To limit the reporting of lower quality emission data, a simple measure of wind persistence was calculated. The wind persistence was calculated by determining the ratio of the mean vector wind speed to the mean scalar wind speed (Eq. 1) (Shirvaikar, 1972). Wind persistence therefore varied between 0 and 1, with a wind persistence value of 1

170   indicating extremely stable winds, and a wind persistence value of 0 indicating extremely unstable (or highly variable) winds. In principle, wind persistence can be calculated for any time averaging period but was calculated here for hourly periods to match the averaging period for the meteorology and CH$_4$ mole fraction data.

$$Wind\ persistence = \frac{|\bar{z}|}{\bar{z}} \tag{1}$$

175

Where $|\bar{z}|$ is the mean vector wind speed, and $\bar{z}$ is the mean scalar wind speed, calculated via Eq. 2 and Eq. 3 respectively.

$$|\bar{z}| = \sqrt{\bar{u}^2 + \bar{v}^2} \tag{2}$$

$$\bar{z} = \overline{\sqrt{u^2 + v^2}} \tag{3}$$

180

Where u and v are the horizontal and vertical components of the wind respectively. The impact of wind persistence on the emissions quantified using different dispersion models was tested. The application of various thresholds for invalidating periods of low wind persistence was also tested.

2.3 Validation setup

185   NPL developed and operate several controlled release facilities (CRFs; IGI Systems Ltd.) capable of releasing multiple gases simultaneously from different release points to mimic 'real-world' emission scenarios. A detailed description of one such CRF is provided by Gardiner et al. (2017) although it should be noted that the CRF operated by Gardiner et al. was a different instrument to the CRF operated as part of this work. The CRFs can be, and have been, used for the validation of a range of gas detection and emission rate quantification solutions (Gardiner et al., 2017; Shah et al., 2019; EN17628:2022, 2022; Defratyka

190   et al., 2023). In brief, the CRFs comprise an array of gas cylinders connected to a network of mass flow controllers (MFCs) calibrated for flow rates across multiple orders of magnitude. The gas release point(s) can be located within approximately 50 m of the CRF equipment. A wide range of accurate release rates can be achieved by using different combinations of MFCs. The CRF used in this work provided accurate and precise releases of different gases; for CH_4, the CRF was capable of volumetric releases from 5 mL min^{-1} to 500 L min^{-1}, equivalent to approximately 0.0002 kg h^{-1} to 21 kg h^{-1} with low uncertainty

195   and metrological traceability (to the SI, or International System of Units). Only one release rate was tested during this validation campaign but there are plans to validate the FEDS system against a range of release rates in future campaigns.

The FEDS validation campaign was conducted over three weeks in March and April 2022. The CRF and FEDS systems were set up in a flat and open area, with no substantial structures or changes in elevation for more than 250 m in all directions, at Bedford Aerodrome (Bedfordshire, UK). For the validation campaign, 12 inlet locations were arranged in a

200   circular pattern around the centrally-located emission release point. Table 2 presents some of the pertinent details regarding the validation campaign, Table 3 some data regarding each release, and Figure 2 illustrates the CRF CH_4 release rates and the FEDS setup. For each release, methane was released continuously (day and night) for at least 20 hours; this temporal release pattern may be representative of continuous emission sources but less representative of short-term (transient) emission sources or emission sources which vary over time.

205

Table 2. Validation campaign details.

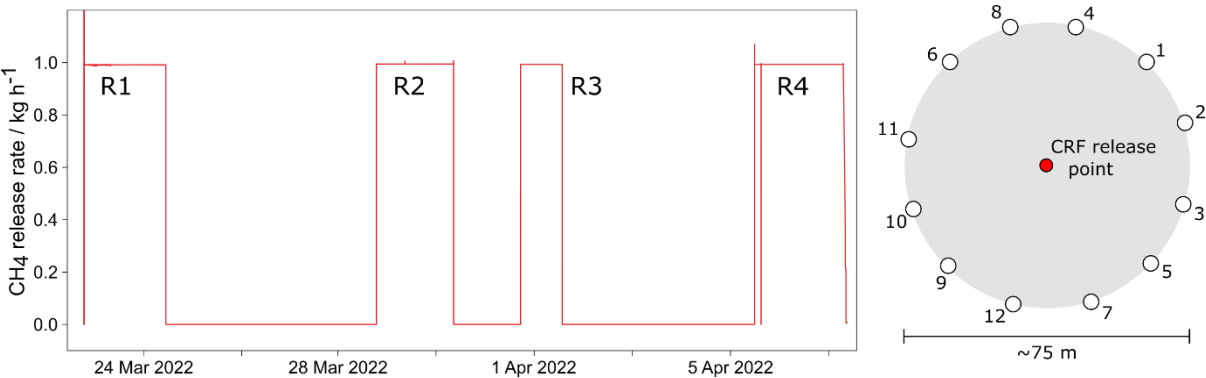| Campaign	Validation 2022
Date	22nd March – 8th April 2022
Location	Bedford Aerodrome
Number of sampling inlets	12
Sampling inlet heights	Fixed at ~2 m
Number of controlled releases	4
Duration of controlled releases	>20 hours
CRF CH_4 release heights	Fixed at ~2 m
CRF CH_4 release rates	1 kg h^{-1}

Figure 2. CRF controlled release rates (left) for each of the four numbered controlled releases (R1 through R4) and FEDS inlet locations (right; numbered 1-12) relative to the CRF release point location (red) for the campaign (see Appendix A for photograph reference). See Table 3 for the mean CH_4 emission rate during each of the four releases. Note that inlet location numbers were primarily for the purposes of tracking metadata and were not indicative of sequential sampling order by the gas analyser (i.e., the sampling order was not in numerical order). Equivalent flows were achieved at each sampling inlet using different vacuum pumps.

2.4 Emission quantification

The spatial scales involved in this work, and for which FEDS is intended (i.e., local and facility-scale) necessitates that any models used for quantifying emissions are capable of simulating atmospheric dispersion over these relatively small spatial scales (<1 km). Many popular inverse models may be more typically suited to modelling atmospheric dispersion over larger spatial scales (>1 km; e.g., the Numerical Atmospheric-dispersion Modelling Environment, NAME, or the FLEXible PARTicle dispersion model, FLEXPART) and were not tested here (Jones et al., 2007; Pisso et al., 2019). Three inverse dispersion models which have been used for facility-scale emissions quantification were used to quantify CH_4 emission rates here. The three models included:

9

- Airviro (v4.0; Apertum, www.airviro.com), an air quality management system with an inverse Gaussian plume dispersion model (Wickström et al., 2010).

225

- WindTrax 2.0 (ThunderBeach Scientific, www.thunderbeachscientific.com), a Lagrangian stochastic particle model used for modelling atmospheric transport over small horizontal distances (<1 km) (Flesch et al., 1995; 2009).

- A custom-built Gaussian plume model, similar in concept to that described in OTM-33A (US EPA, 2014).

230   The three models, and their set up, are described in more detail in Appendix B.

Two different approaches to quantifying emissions were tested. Firstly, emissions were quantified using knowledge of the precise CH_4 controlled release timings. This was done by creating subsets of CH_4 and meteorological data from which emissions could be quantified. This approach required knowledge of emission event timings to be known; something which is unlikely in real-world scenarios. The second approach assumed no prior knowledge of controlled release timings and is

235   therefore likely to be a better reproduction of the real world, where exact knowledge of the timings of emission events is unlikely. For this second method, data were analysed in temporal intervals of varying lengths (e.g., daily [24-hourly], 12-hourly, 6-hourly etc.).

Inverse dispersion models are known to be sensitive to many user-defined parameters and inputs. The sensitivity of estimated emission rates to some parameters (e.g., release height, sensor height, CH_4 atmospheric background) was tested to

240   validate the performance of the FEDS and the subsequent approaches to emission quantification. Further work on understanding the sensitivity of reverse dispersion models to multiple variables will be undertaken in the future.

3 Results and discussion

This section presents and discusses results from the validation of the FEDS method.

3.1 Meteorology

Figure 3 shows hourly-averaged wind conditions (wind speed, wind direction, standard deviation in wind direction (1σ), and wind persistence) measured over the course of the validation campaign.
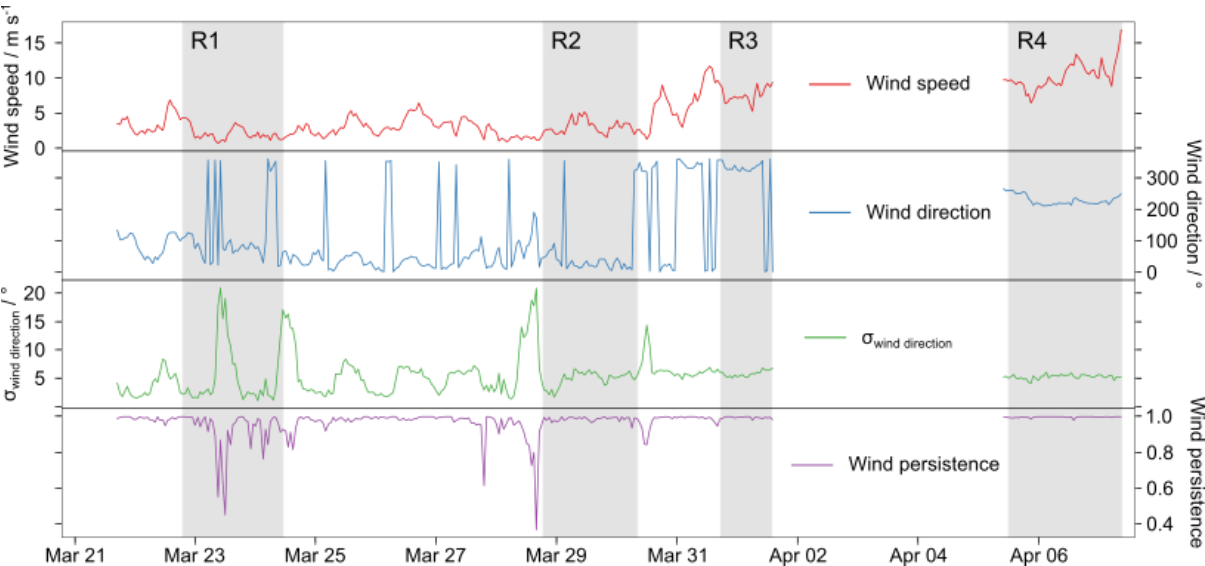

Figure 3. Hourly-averaged wind speed, wind direction, standard deviation (1σ) in wind direction (σ_wind direction), and wind persistence (see Eq. 1) measured at 10 m above ground from the hydraulic mast on the FEDS trailer. Meteorology was not measured between 2nd and 5th April as the trailer was shut down. The four controlled release periods (R1 through R4) are shown in grey.

Wind speed was generally below 5 m s^{-1} before 30th March and above 5 m s^{-1} after 30th March. The wind direction before 30th March was generally northeasterly, switching to northerly between 30th March and 2nd April, and then south-westerly on 5th and 6th April. Figure 4a shows a wind rose for the campaign. Periods of high standard deviation in wind direction occurred when wind direction was highly variable across the hour averaging period. Periods of lower wind persistence (e.g., <0.9) were typically associated with periods of higher standard deviation in wind direction, as expected. All occurrences of wind persistence below 0.9 were associated with average wind speeds of less than 2.5 m s^{-1}. Higher wind speeds (> 5 m s^{-1}) were almost always concurrent with wind persistence greater than 0.95. Mean hourly wind persistence (and range in wind persistence) across each of the four controlled releases is provided in Table 3.

3.2 Methane mole fraction results

Figure 4 shows a broad picture of CH$_4$ mole fractions measured from each sampling inlet location under differing wind directions. The CH$_4$ rose plots are arranged around the CRF release point location corresponding to the relative sampling

11

location (see Figure 2). The wind rose plot (also shown on Fig. 4) indicates the dominance of certain wind directions (northerly and northeasterly) during the campaign. Higher CH_4 mole fractions (>3 ppm) were measured at inlets located to the south and

265   southwest of the CRF (i.e., inlets 7, 9, and 12) corresponding to those locations downwind of the CRF during northerly and northeasterly winds. High CH_4 mole fractions were also measured at inlet 1 (located to the northeast of the CRF) during southwesterly winds, and at inlets 10 and 11 (located to the west of the CRF) during easterly winds. Those inlet locations which were not downwind of the CRF generally measured lower CH_4 mole fractions (<2.5 ppm) during northerly and northeasterly winds. These measurements demonstrate that the controlled release of CH_4 was correctly sampled by the FEDS as it cycled

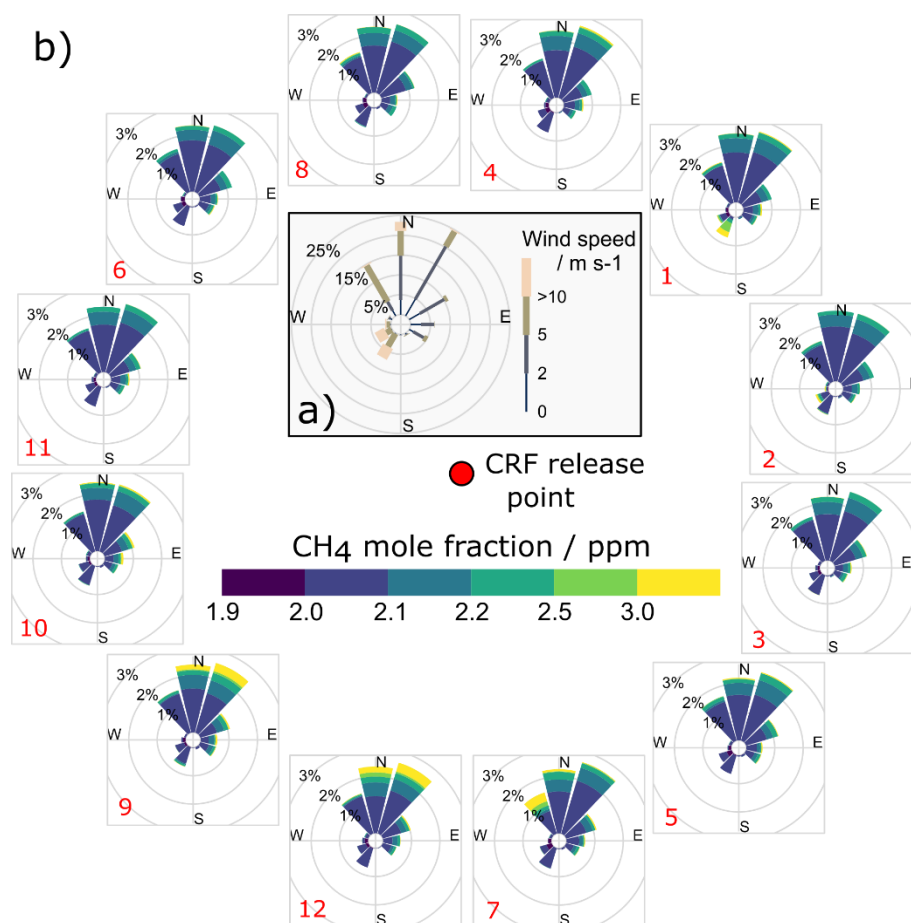270   through sampling locations.

Figure 4. a) Wind rose showing the frequency of wind speeds in 12 wind direction bins for the whole campaign period, and b) CH_4 rose plots showing CH_4 mole fractions measured at each inlet (arranged around the central CRF release point location as shown in
275   **Figure 2) during the campaign period. The radii of the wedges in 12 wind directions shows the frequency of mole fractions occurring in the coloured mole fraction bins. The red text indicates the inlet number corresponding to the CH_4 mole fraction rose.**

3.3 Quantifying emissions using prior knowledge of release timings

In this section, we assess the performance of the dispersion models for quantifying CH_4 emissions using prior knowledge of the timings of the controlled releases of CH_4. In this way, the data supplied to, or output from, the models was directly related

280 to the periods of time in which CH_4 was being released. For Airviro, the model time step was set to cover the release period for each individual release. For WindTrax, the hours of output data corresponding to release periods were averaged. For the Gaussian plume analysis, only data within the release periods were considered.

Table 3 shows the mean controlled release rates from the four releases, and the emission rates quantified using the different dispersion models. The wind persistence (mean of hourly wind persistence values) for the duration of each controlled

285 release is also provided.

Table 3. Controlled (mean) CH_4 release rates, wind conditions, and quantified emission rates (Airviro[a], WindTrax[b], and Gaussian plume[c]) for each of the four controlled releases. A value of NA indicates that no emission rate was quantified by that model. The quoted uncertainties for the wind persistence, wind speed, and controlled release rates are 1 standard deviation (1σ) about the mean

290 **across the release period. The model error output by Airviro is described as a standard error, defined as the standard deviation of the estimated error in Wickström et al. (2010). The model error output quoted here for WindTrax is the standard deviation (1σ) in individual hourly emission rates over each release period. The uncertainty value for the Gaussian plume model is the combined uncertainty in the Gaussian plume equation (see Appendix B).**

Release rate number	Approximate duration / hours	Wind persistence (*range*)	Wind speed / m s⁻¹	Controlled release rate / kg h⁻¹	Airviro [a] / kg h⁻¹	WindTrax [b] / kg h⁻¹	Gaussian plume [c] / kg h⁻¹
R1	40	0.92 ± 0.12 (*0.45 – 1.00*)	2.0 ± 1.0	0.991 ± 0.011	NA	1.96 ± 6.20	0.30 ± 0.19
R2	36	0.99 ± 0.01 (*0.95 – 1.00*)	3.1 ± 1.1	0.994 ± 0.011	0.79 ± 0.12	0.82 ± 1.05	0.69 ± 0.36
R3	21	0.99 ± 0.01 (*0.94 – 1.00*)	7.4 ± 1.5	0.993 ± 0.011	1.08 ± 0.16	1.72 ± 1.89	0.95 ± 0.41
R4	43	0.996 ± 0.004 (*0.98 – 1.00*)	10 ± 1.9	0.969 ± 0.012	1.08 ± 0.09	0.99 ± 0.70	0.91 ± 0.38

[a] Airviro model settings (see Appendix B): sampling lag time = 90 s; release height = 2 m; emission zone = 3 (see Appendix C); CH_4

295 background value = internally calculated.

[b] WindTrax model settings: sampling lag time = 90 s; release height = 2 m; point source emission; particle number = 50,000; CH_4 background value = 2.00 ppm.

[c] Gaussian plume model settings: sampling lag time = 90s; wind direction bin width = 2°; CH_4 background = 5th percentile of data within release window (1.99 – 2.03 ppm); median stability class.

300

All three models estimated emission rates within a factor of two of the controlled release rates, except during R1. Airviro was unable to output an emission rate for R1, whilst the Gaussian plume model estimated an emission rate of 0.30 ($\pm$ 0.19) kg h^{-1}. WindTrax also performed poorly for R1, with a mean estimated emission rate of 1.96 ($\pm$ 6.20) kg h^{-1}. This may be explained by the wind persistence during R1, which was much lower (0.92) than for the other three release experiments (0.99, 0.99, 0.996

305   for R2, R3, and R4 respectively). The higher wind persistence (0.996) observed for R4 correlated with better performance from all three models, with estimates of 1.08 ($\pm$ 0.09), 0.99 ($\pm$ 0.70), and 0.91 ($\pm$ 0.38) kg h^{-1} for Airviro, WindTrax, and the Gaussian plume simulation respectively.

Whilst the mean emission rates estimated by WindTrax were within a factor of two of the controlled release rates, the individual hourly emission rates varied considerably over each of the release periods. For example, hourly estimates of

310   emission rates for R1 ranged between 0.006 and 33.4 kg CH$_4$ h^{-1}, thus explaining the large standard deviation about the mean for R1. If shorter controlled releases (on the order of a few hours) were used, any results may therefore have been subject to large absolute biases. However, it should be noted that the WindTrax setup used in this study was not optimal; WindTrax typically requires parity between the number of emission sources and the number of concentration measurement locations (i.e., one concentration measurement for each emission source). The design of the FEDS validation campaign meant that this was

315   not possible and therefore WindTrax was subject to increased model uncertainty. Furthermore, a default value for atmospheric turbulence was used for all WindTrax simulations. Catering WindTrax model simulations to conditions at the time of each release, for example by using different Pasquill stability classes, may have improved results. A single, constant value for turbulence may also have precluded the ability for WindTrax to model diurnal variations in atmospheric stability. The use of measured temperature data in Airviro may have aided that model's treatment of atmospheric turbulence and stability.

320   The source location (SL) module in Airviro (see Appendix B) also has the capability to estimate the likely emission source location(s) based on the input data. Figure 5 shows the probability of emission source locations output by Airviro. It is clear that for R1, Airvrio was incapable of correctly locating the emission source location. This is potentially also correlated with the lower wind persistence during R1 (0.92 $\pm$ 0.12), and likely explains the lack of an output emission rate for that release (Table 3). For R2, R3, and R4, Airviro was able to estimate the location of the emission source, albeit with greater confidence

325   during R4 than R3, and during R3 than during R2. The increased confidence with locating the emission source correlated with higher wind persistence, as well as with reduced percentage uncertainty in the emission rate. This implies that Airviro performs better under more consistent wind conditions, and that emission source locations and quantified emission rates under inconsistent wind conditions may be subject to greater uncertainty.
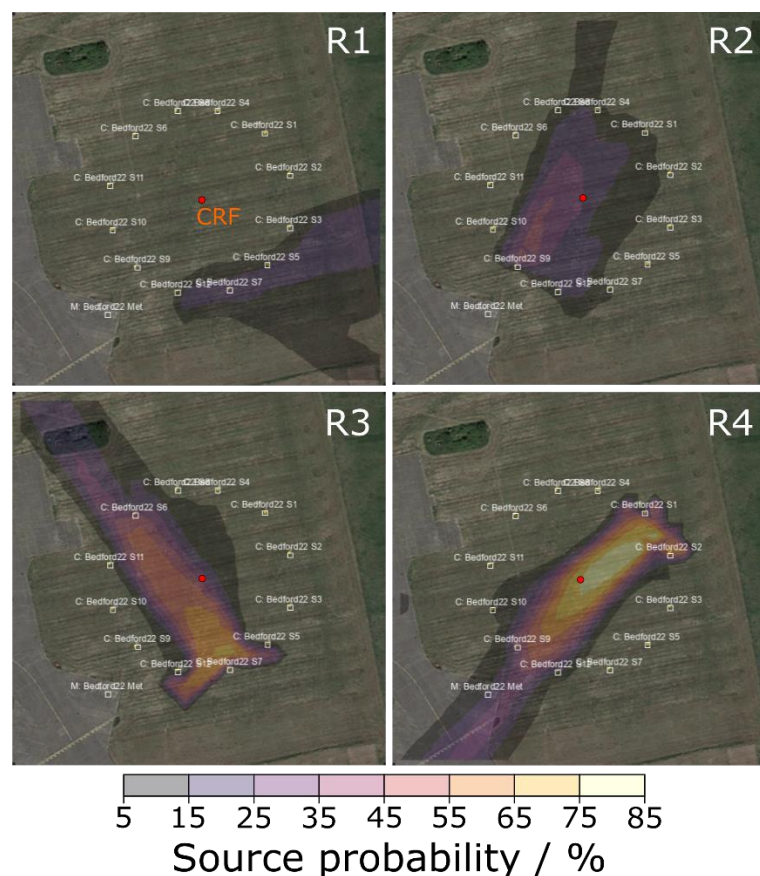
Figure 5. Airviro output 'source probability' for the four release periods (R1, R2, R3, and R4). The coloured contours show the probability that the emission source is located within each contour shown by the colour scale. The red dot shows the true location of the controlled release facility (CRF) from which CH$_4$ was released. Satellite imagery: Google, ©2020 Maxar Technologies.

Figure 6 shows the Gaussian plume functions for each of the four controlled releases, following binning of the CH$_4$ data into 2° wind direction bins. The correlation between CH$_4$ enhancements and the wind direction blowing directly towards the sampling inlets (i.e., rotated wind direction closer to zero) was clear in the case of R2, R3, and R4. However, there was little correlation between CH$_4$ enhancement and wind direction for R1. Despite this, a Gaussian curve was still generated, and the method estimated a final emission of 0.30 ± 0.19 kg h^{-1}. Emission rates were estimated using the highest point of the Gaussian curve (a1; see Appendix B). Whilst the result for R1 was lower than the controlled release rate, it was still within an order of magnitude, though this may be a coincidence rather than proof that this method works even under poor wind conditions. The spread of CH$_4$ enhancement data for the other three releases was much tighter, clearly demonstrating that enhancements were only measured at sampling locations when the wind was blowing towards them (see Fig. 4). As was the case for the other two dispersion models, the Gaussian plume estimated emission rate and the calculated uncertainty appeared to be correlated with the wind persistence for the duration of each controlled release.

15

345   It should also be noted that averaging concentration data into wind direction bins over long periods of time may combine CH_4 enhancements measured under very different wind speed conditions. This may explain the high variability in CH_4 data measured even within the plotted Gaussian function (see grey data in Fig. 6). This simple Gaussian plume approach may therefore be further refined by ensuring that the wind field, including the wind speed, is consistent for the duration of any analysed time periods.
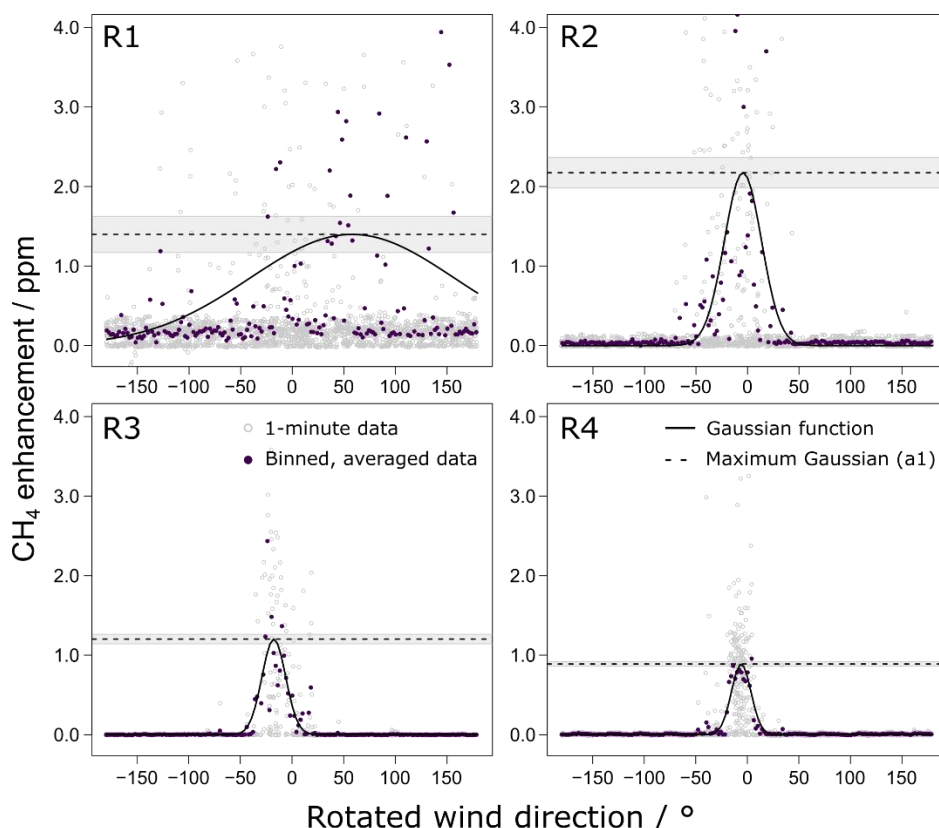
350

Figure 6. OTM-33A analysis and fitted Gaussian plume function for the four controlled releases, using a bin width of 2° and a CH_4 background value equal to the 5th percentile of CH_4 data during each release. The maximum CH_4 enhancement was determined by the maximum point of the Gaussian function (see Appendix B), and is denoted here by the dashed black line, with the uncertainty
355 in this value shown in grey.

3.4 Quantifying emissions without prior knowledge of release timings

In this section, the performance of the reverse dispersion models for quantifying emissions without prior knowledge of the release timings is assessed. This approach more closely mimics a real-world scenario, in which precise (or even approximate)
360 knowledge of emission event timings may be unknown. The impact of changing different model variables and parameters on model performance and sensitivity is also examined in this section. Only Airviro and WindTrax are analysed, as the custom

16

Gaussian plume model (based on OTM-33) has less scope for varying parameters. Some additional model parameters are also analysed in Appendix C but will be the focus of future work assessing the sensitivity of dispersion models to multiple variables.

3.4.1 Changing model time step and averaging period

365 In the absence of information regarding the timings of emission events, it is expected (and plausible) that dispersion models will be used to evaluate data over regular periods of time. Airviro has scope for setting the time period (via a start and end time for a simulation), corresponding to multiple time intervals of input data, over which data is evaluated. WindTrax only outputs emission rates at intervals equal to the timestep of input data (in this case, hourly). Whilst this does not allow for equivalent comparison between the two models, the 1-hourly data input to WindTrax was averaged over longer periods of time (e.g.,

370 daily) to allow for indirect comparison to the changing Airviro timestep length. Time averaging for WindTrax was therefore applied to both the input data (concentration and meteorology) and to the output data (emission rate).
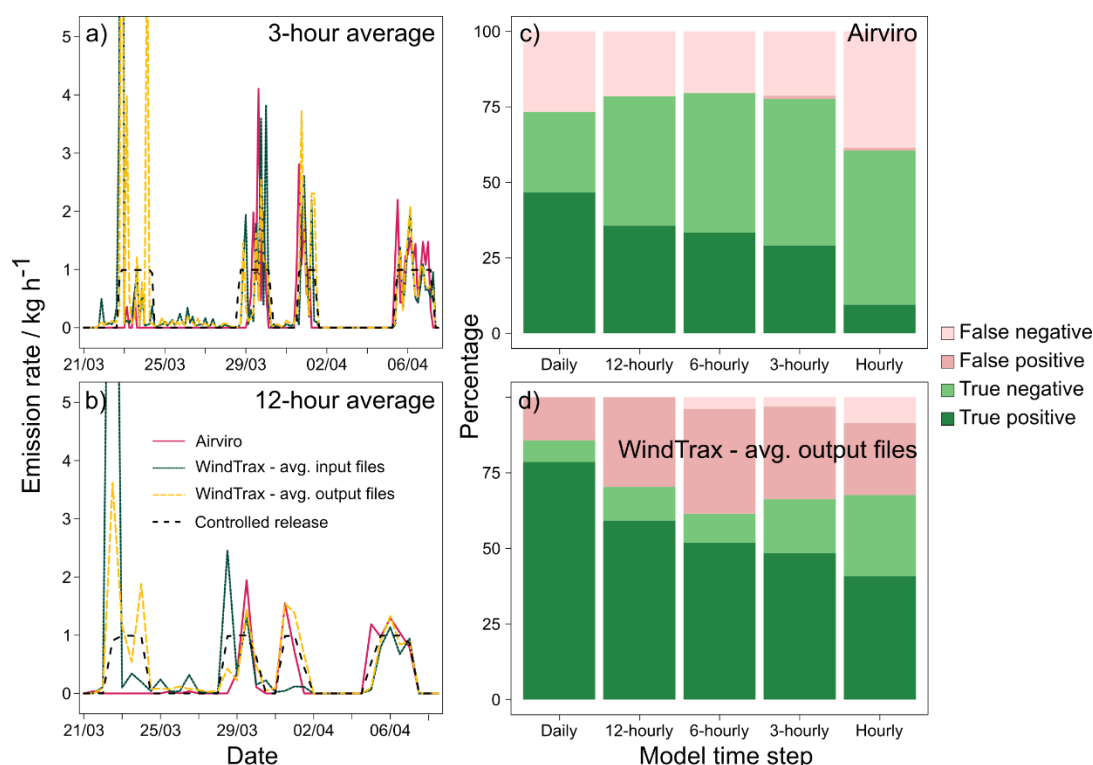
  Figure 7 shows time series of Airviro and WindTrax emission rate estimates (where emission rates were calculated either by time-averaging hourly input data prior to model evaluation, or by time-averaging hourly output data following model evaluation) compared with CRF release rates, for 3-hourly and 12-hourly periods (see Appendix C for time series of other

375 averaging periods). Whilst the average emission rate estimated by the models for the duration of a controlled release was broadly similar with the release rate (Table 3), there was a lot of variability in the emission rates estimated by both models over shorter time periods. This variability was far more evident for shorter averaging periods (i.e., in the 3-hour average emission rates compared with the 12-hour emission rates). For example, WindTrax estimated emission rates ranged between 0.003 and 33.4 kg h^{-1} on an hourly basis. The impact of wind persistence is also visible on these time series, especially for

380 Airviro. Airviro failed to estimate any emissions during the first controlled release (R1; wind persistence 0.92).

  The capacity of the models for correctly capturing emission events (and periods without emissions) can be evaluated using contingency metrics (Bennett et al., 2013). The righthand panels in Figure 7 show the percentage of true positives (in which the model correctly identifies emissions), false positives (in which the model incorrectly identifies emissions), true negatives (in which the model correctly identifies periods of no emissions), and false negatives (in which the model incorrectly

385 identifies periods of no emissions) for both Airviro and WindTrax using different timesteps. For this analysis, an arbitrary threshold of 0.05 kg h^{-1} was chosen, above which output emission rates were considered an event, and below which output emission rates were considered a non-event. A threshold was selected to exclude negligible emission rates (e.g., 0.003 kg h^{-1}) from counting as an emission event detection. Airviro showed a large proportion of false negatives in all time averaging cases; false negatives were generally associated with R1 and therefore potentially related to the lower wind persistence during that

390 release. WindTrax had fewer false negatives (emissions were correctly identified during R1 even if the magnitude wasn't well constrained) but a greater proportion of false positives, where emissions above the arbitrary threshold (>0.05 kg h^{-1}) were estimated when no controlled releases were taking place. It should be noted that the lack of meteorological data between R3 and R4 meant that model contingency during this period was not evaluated. Airviro had a greater percentage of correct results (true positives and negatives) for 12-hourly and 6-hourly timesteps, whereas WindTrax had the greatest proportion of correct

395   results when using daily averaging, and the lowest proportion of correct results when using 6-hourly averaging. It should be
        noted that the relative proportion of time spent with and without controlled releases of emissions may bias the results in favour
        of one or other of the models; a model which never output an emission rate would still yield true negatives some of the time
        whilst a model which always modelled emission rates greater than zero would give true positives. The time period when the
        FEDS trailer was shut down (2nd-5th April) was not analysed here (see Fig. 3).

400

Figure 7. a) Time series of 3-hour average emission rate results for Airviro and WindTrax (averaging either the input data or 1-hourly output results for WindTrax). The black dashed line shows 3-hour average controlled release rates. b) Time series of 12-hour average emission rate results for Airviro and WindTrax (averaging either the input data or 1-hourly output results for WindTrax).
405   **The black dashed line shows 12-hour average controlled release rates. c) Contingency matrix (event observation) for Airviro with different time steps. d) Contingency matrix (event observation) for WindTrax with different averaging of hourly output emission rates. All results shown with a lag time of 60 seconds and a plume height of 2 m.**

        The success index provides a measure of the ability of a model to correctly detect occurrences and non-occurrences of
410   emissions (Bennett et al., 2013). Table 4 gives the success index for Airviro and WindTrax with different timesteps or
        averaging periods. Airviro with daily, 12-hourly, and 6-hourly timesteps performed the best, despite being unable to detect
        emissions during R1. However, the success of Airviro dropped off with shorter timesteps. WindTrax performed most
        successfully using daily and hourly averaging but did show such a substantial reduction in success rate when using intermediate
        length averaging periods. Airviro had broadly greater success rates than WindTrax despite not estimating emission rates during

415 R1. This was because WindTrax output emission rates (and sometimes substantial emission rates) for all hours of input data regardless of if emissions were truly present or not.

Whilst the success index provides an indication of correct interpretation of event timings, the success index does not provide information on the performance of the models for accurately estimating the rates of emission. Table 4 also provides a bias score (or mean error) and root mean square error (RMSE) for each of the model iterations. Airviro underestimated

420 emissions on average (negative bias), though this may be exaggerated by the lack of estimated emissions during R1. WindTrax, on the other hand, overestimated emissions on average (positive bias). The RMSE was generally lower for Airviro than for WindTrax, largely due to the vast overestimation of some emission rates during R1 by WindTrax (especially when using shorter time averaging). WindTrax, with a daily averaging period, performed the best overall, with the lowest RMSE (0.33 kg h^{-1}) and an average bias of -0.02 kg h^{-1}. Airviro with 12-hourly or 6-hourly periods performed next best, with a small negative

425 bias (of -0.15 and -0.11 kg h^{-1} for 12-hourly and 6-hourly respectively), and RMSE scores of 0.51 and 0.53 kg h^{-1} for 12-hourly and 6-hourly respectively. A timestep (or averaging period) of 1-hour performed the worst when compared with other timesteps for both models.

Table 4. Metrics comparing the Airviro and WindTrax modelled emissions rates with the controlled release emission rates whilst
430 **varying the time step (or averaging period).**

Model	Time step (or averaging period)	Success index / %	Bias / kg h^{-1}	RMSE / kg h^{-1}
Airviro	Daily	82	-0.30	0.54
	12-hourly	81	-0.15	0.51
	6-hourly	81	-0.11	0.53
	3-hourly	78	-0.12	0.63
	Hourly	59	-0.32	0.75
WindTrax	Daily	67	-0.02	0.33
	12-hourly	64	0.14	0.62
	6-hourly	57	0.18	1.06
	3-hourly	65	0.17	1.38
	Hourly	68	0.19	2.39

3.4.2 Wind persistence

The impact of low wind persistence (or high wind direction variability) on emission estimation was observed during R1 in Section **Error! Reference source not found.**. Airviro, in particular, struggled to estimate emissions during those releases with l

435 ower average hourly wind persistence. There was a small correlation between the error in estimated emissions and the average hourly wind persistence ($R^2 = 0.58$; see Appendix C), with generally greater error in estimated emission rate during periods of

lower wind persistence. Several thresholds for excluding estimated emission data were trialled, using values of average hourly wind persistence below which emission data for that time period were considered invalid. Results (success index, mean bias, and RMSE) following application of these thresholds are shown in Figure 8, for wind persistence threshold values of: no

440   threshold (equivalent to all data), 0.85, 0.90, and 0.95. The proportion of time periods eliminated by the threshold increased with increasing threshold value.

Using a wind persistence threshold improved the success index for Airviro in all cases but reduced the success index for WindTrax in some cases (see Appendix C for contingency matrices before and after application of a wind persistence threshold). Improvement in success index for Airviro but not for WindTrax was likely due to the different performance of

445   these models during R1. The first release (R1) occurred during a period of low wind persistence and was therefore invalidated when using a wind persistence threshold. As Airviro failed to output an emission rate during R1, removing R1 from the analysis improved the success index for the remaining releases.

Figure 8 clearly demonstrates the negative bias in emission estimates output by Airviro. The bias in Airviro results generally reduced when using any wind persistence threshold (especially with 12-hourly time steps). The positive bias in

450   emission estimates is also evident for WindTrax. For WindTrax, a greater bias was observed when using wind persistence thresholds for 12-hourly and 6-hourly averaging, but smaller biases were observed when using wind persistence thresholds for other time averaging periods. The impact of including wind persistence thresholds on the RMSE was mixed for both models. For Airviro, reductions in RMSE occurred for daily time steps but there was little impact when using shorter time steps. For WindTrax, the RMSE reduced slightly when using a wind persistence threshold of 0.95 with daily time averaging but actually

455   increased when using a wind persistence threshold of 0.95 with 12-hourly averaging. Interestingly, for WindTrax, the RMSE improved when using a wind persistence threshold with the shorter time-averaging periods. Appendix C provides values for success index, bias, and RMSE with wind persistence thresholds of 0.90 and 0.95.

Overall, daily averaging periods appeared to work best for both Airviro and WindTrax, with generally greater success index, and lower values for both bias and RMSE. Slightly shorter time periods (i.e., 12-hourly) appeared to be the next best

460   averaging period. Averaging data over 12-hourly periods also had the benefit of providing a greater number of observations for analysis (28 12-hourly data points compared to 15 daily data points). The results whilst testing different wind persistence thresholds were more mixed between Airviro and WindTrax but a threshold of 0.95 had the lowest RMSE for both models under both daily and 12-hourly averaging. For these reasons, all experimental results following are demonstrated for 12-hourly averaging with a wind persistence threshold of 0.95.
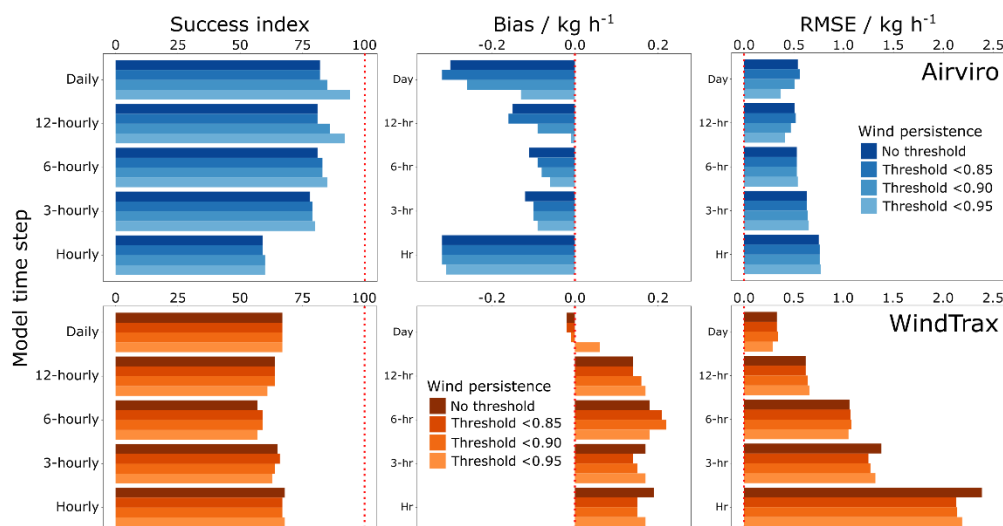
465

Figure 8. Graphical summary of success index (left column), bias (central column), and RMSE (right column) for Airviro (top row; blue) and WindTrax (bottom row; orange). Ideal values are indicated by red dashed lines (success index = 100%; bias = 0 kg h^{-1}; RMSE = 0 kg h^{-1}). Different shaded colours represent results with different average hourly wind persistence thresholds, where data was removed if the wind persistence was below a certain value.

3.4.3 Changing number of sampling locations

The number of sampling locations required to accurately monitor an emission source is a key parameter for the FEDS system and any other system with a distributed sampling network. The more locations installed, the more equipment and resources required and the longer the time taken to deploy that equipment. A minimum number of sampling locations would be preferential in most cases to reduce equipment costs and labour.

Figure 9 shows the RMSE (kg h^{-1}) results for Airviro and WindTrax simulations with data input from differing numbers of sampling locations. A clear trend was observed with RMSE decreasing in magnitude with increasing numbers of sampling locations. A similar trend was observed for the success index (data point colours), with model simulations more successful at correctly identifying emission events when data from more sampling locations were used. The exact configuration of locations, along with RMSE and success index scores, can be found in Appendix C.

Interestingly, the RMSE was lowest for WindTrax when using data from three sampling locations. Whilst the success index was also low (and therefore quite poor) for these simulations, the low RMSE values may be the result of the WindTrax model's optimisation procedures. WindTrax seeks to resolve a least-squares statistical fit to the data based on the geometric arrangement of sensors and sources; having more emission sources than concentration sensors can lead to an overdetermined system of equations and thus greater uncertainty (Crenna et al., 2008).
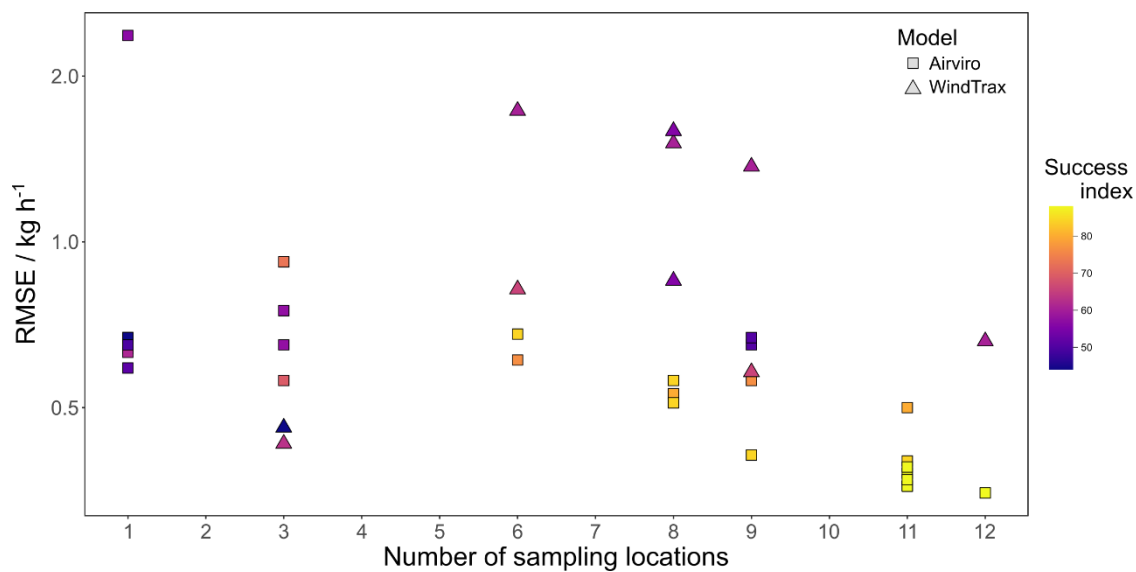
Figure 9. RMSE results as a function of the number of sampling locations for both Airviro (square data points) and WindTrax (triangular data points). Data points are coloured by the success index.

490

4 Conclusions and future improvements

NPL's Fugitive Emission Distributed Sampling (FEDS) system is a mobile, multi-inlet sampling system. FEDS has important advantages over other network systems in that only a single high-performance trace gas analyser is required, reducing the cost

495   relative to systems using multiple high-performance instrumentation. Controlled releases of CH_4 were used to validate the FEDS system for measuring and estimating CH_4 emission rates from point sources on the facility scale. Higher concentrations of CH_4 were measured when sampling from locations downwind of the controlled release source. Data were collected from each individual sampling location for only a few minutes each hour, but hourly-averaged data was suitable for input to reverse dispersion models for emission source location and emission rate estimation.

500   The performance of three reverse dispersion models for the estimation of emission rates was tested. Individual hourly emission estimates from all models had high error (high RMSE) when compared to controlled release rates, but estimates were greatly improved when averaged over time periods greater than 12-hours (lower RMSE). The performance of all three models was also shown to be closely correlated to the wind conditions, with a highly variable wind field (assessed via wind persistence) generating circumstances in which the models performed less well, or with greater uncertainty. Higher wind persistence (or

505   low variability in wind direction) during release periods improved performance. This suggests that the wind direction variability must be low for the course of a full measurement cycle (measuring from all sampling locations).

The performance of the models was also improved when using prior knowledge of the timings of the emission event. However, it is acknowledged that this information may be unknown in real-world scenarios. This work provides evidence that Airviro and WindTrax generally worked well in the absence of knowledge of the timings of releases; emission rates were

510   quantified reasonably accurately (low bias, low RMSE) using daily and 12-hourly averaging of input data (CH_4 concentration and meteorology), especially for Airviro, and especially after removing periods of lower wind persistence below a threshold of 0.95. For daily averaging periods, the RMSE was 0.37 kg h^{-1} and 0.29 kg h^{-1} for Airviro and WindTrax respectively whereas when using hourly averaging periods, the RMSE was 0.77 kg h^{-1} and 2.19 kg h^{-1} for Airviro and WindTrax respectively.

Of the three reverse dispersion models tested, Airviro performed the best and most consistently, as well as offering

515   additional advantages in being able to provide estimates of the emission source location. WindTrax, on the other hand, required the emission source location to be specified before emissions could be quantified. Alternative models capable of modelling transport on small spatial scales (<1 km) exist, and FEDS could be used to validate these models for estimating CH_4 emissions. Further work is also required to test the sensitivity of these dispersion models (and output emission rates) to user-specified input parameters, such as the CH_4 background concentration or particle number. Some work to this effect is covered in

520   Appendix C but is not covered here in great detail.

The use of only a single trace gas analyser does come at the cost of reduced temporal coverage at each sampling location. Only four minutes of data were collected at each sampling location every hour (with at least the first 60 seconds of data discarded after switching inlets in case of contamination in the optical cell due to slow cell flushing time). This remains the major disadvantage of using only a single analyser measuring from multiple locations over using multiple analysers in a

525   FEDS-type system. However, it should be noted that the use of multiple analysers may introduce an element of inter-instrument variability to the data. Such variability may be mitigated with robust calibration procedures but may not be ruled out, especially if analysers are of different types and/or manufacturers. In the absence of a direct comparison between using one and using multiple analysers, it is difficult to conclusively determine if measurement quality is proportionately improved when using a more expensive system with multiple analysers.

530   Low-cost sensors were also deployed alongside the single high-performance gas analyser for the measurement of CH_4 at multiple sampling locations. Data from the low-cost sensors were unsatisfactory for accurate measurement of CH_4 and robust calibration of the data remains difficult at this point in time. Future FEDS development will continue to trial the incorporation of low-cost sensors for the continuous measurement of CH_4 at each of the sampling locations. This will increase the temporal capture of CH_4 concentration at each location up to 100%, albeit with a potentially reduced data quality (relative

535   to the performance of the analyser). The four minutes of co-located data measured by the gas analyser could be used to correct the low-cost sensor data to improve data quality. However, given current concerns around low-cost sensor reliability for accurate measurement of CH_4, it is not expected that low-cost sensors could be used as a direct substitute for a high-performance gas analyser.

The validation approach here used controlled releases of CH_4 over reasonably long periods of time (>20 hours) and

540   hence the performance of FEDS for transient emission events of considerably shorter duration (i.e., hours or minutes) was not assessed. Future validation experiments could seek to examine the impact of other variables on emission quantification performance including the magnitude of emissions (continuous or variable), release duration, sensor or receptor height(s), topography, and obstructions to dispersion.

545

Appendices

Appendix A: Additional contextual information

Table A1 shows the proportion of time spent in each Pasquill stability class during each of the four controlled releases. The Pasquill stability class was calculated using the method prescribed in OTM-33A (US EPA, 2014) from the measured standard

550 deviation in wind direction.

Table A1. Proportion of time (1-minute average) spent in each Pasquill stability class for each of the four controlled releases.

Release rate number	Stability class / %						
	1 *Extremely unstable*	2 *Moderately unstable*	3 *Slightly unstable*	4 *Neutral*	5 *Slightly stable*	6 *Moderately stable*	7 *Extremely stable*
R1	1.7	0.7	1.4	1.9	3.8	8.6	81.8
R2	0.0	0.0	0.0	0.0	0.4	8.6	91.0
R3	0.0	0.0	0.0	0.0	0.2	8.1	91.7
R4	0.0	0.0	0.0	0.0	0.1	5.6	94.3

555

Figure A1. Aerial photograph of the validation setup (see also Fig. 2). The FEDS trailer can be seen in the centre foreground of the image (behind the van) with the 10 m meteorological mast attached to the trailer. The 12 sampling locations are arrayed in a circle on the grass beyond the trailer with the CRF release location in the centre.

560 **Appendix B: Emission quantification models**

This appendix describes the three models used for emission quantification.

Airviro

Airviro (v4.0) is an air quality management system which contains a number of dispersion modelling tools (Apertum, www.airviro.com). Airviro includes an inverse Gaussian plume dispersion model referred to as the Airviro Receptor model

565 (Wickström et al., 2010). The Airviro Receptor model comprises two distinct modules; the source location (SL) module and the emission estimation (EE) module. These two modules can be run sequentially or separately. The Airviro Receptor model uses time series data and can be run across any time period (greater than the time resolution of the input data) for which data is available. There must be data in the Airviro model for all receptors (i.e., sampling locations for FEDS) and meteorological stations utilised for each time step being modelled in order for the model to complete. As CH_4 concentration data were

570 measured sequentially (and not continuously) at each sampling location, meteorological and CH_4 concentration data were both input to Airviro as hourly averages (separate data for each sampling inlet location in the case of CH_4). The 'timestep duration' in Airviro (e.g., hourly, 12-hourly, daily) was specified directly via a start and end time for each simulation, allowing for the estimation of emissions over different time scales. Vertical dispersion in Airviro was modelled using input temperature data from two different heights.

575 Source Location (SL) module

The SL module is used to estimate the emission source location and outputs a map highlighting the regions of highest 'source probability' in which Airviro estimates the greatest likelihood of an emissions source(s). The SL module utilises a user defined grid and cell resolution (typically 10 m × 10 m). Initially, the model places a "pseudo-source" at the centre of each grid cell and runs the Gaussian model forwards to calculate expected concentrations at various receptors (i.e., sampling locations). For

580 each "pseudo-source", the correlation between the simulated and the measured concentration-time series is calculated. The calculations are repeated for all pseudo-sources (one for each grid cell). The horizontal distribution of the correlation coefficients is displayed as isolines multiplied by 100 to yield a 'source probability' at each location. Information from the SL module may then be used by the user to constrain the EE module of the Airviro Receptor model.

Emission estimation module

585 The EE module requires the user to specify area source(s) (or emission zones) as rectangular boxes. The user also specifies an associated source height for each emission zone. The EE module assumes a constant background concentration of the target species (in this case CH_4). Within the EE module, Airviro is constrained to only attribute emissions to the user-specified emission zone(s) within the model grid (unlike in the SL module). Following input of emission zones, Airviro divides each emission zone area into nine "sub-areas". Initially, the EE module assumes each sub-area has a constant, default emission and

590 runs the dispersion model forwards to calculate the expected concentrations at the receptors (as in the SL module). The EE module performs a multiple regression analysis to simulate source contributions and optimises this with the actual measured

concentration data. Airviro outputs a best estimate of the emission rate for each emission zone, along with an associated error. Whilst the error associated with the emission estimated by Airviro is a useful parameter, it should not be considered equivalent to a full uncertainty budget. Firstly, the model error does not account for uncertainty in the measured input parameters (either

595   concentration or meteorology). Secondly, inverse dispersion modelling of time-series data where sources are not always well understood is a highly complex and indirect method of deriving emission rates and therefore may be expected to have a reasonably high associated uncertainty (potentially up to 50% of estimated emissions but likely highly dependent on the quality of measurements and environmental conditions at the time of monitoring, as well as the fidelity of the model). The errors in emission rates output by Airviro are frequently less than 10% of the estimated emission rate. This may be unrealistically low.

600   NPL is investigating the development of a robust uncertainty budget for local-scale reverse dispersion models, such as the Airviro Receptor model.

WindTrax

WindTrax 2.0 (ThunderBeach Scientific, www.thunderbeachscientific.com) is a Lagrangian stochastic particle model used for modelling atmospheric transport over small horizontal distances (<1 km) (Flesch et al., 1995; 2009). WindTrax is specified in

605   the United States Environmental Protection Agency (EPA) Other Test Method 33 (OTM-33) (US EPA, 2014). OTM-33 details a standardised method for the geospatial measurement of air pollutants and remote emissions quantification, typically applied to monitoring from a mobile platform. For context, Other Test Methods are those methods which are not yet subject to the US Federal rulemaking process but may still be used in Federally enforceable programs. WindTrax has been used to quantify CH_4 emissions from landfill and hydraulic fracturing (e.g., Riddick et al., 2017; Shaw et al., 2020).

610   Here, the CH_4 emission rate was estimated using the modelled advection of 50,0000 particles from the release location. A surface roughness length (Z_0) of 2.3 cm (equivalent to short grass) was used. Different particle numbers were examined for the impact on estimated emission rate, but different surface roughness lengths were not tested. Meteorological and CH_4 concentration data were input to WindTrax as hourly averages (separately for each sampling location in the case of CH_4). Atmospheric turbulence was modelled using a default value for the Monin-Obukhov length. The WindTrax timestep

615   cannot be specified and is always equal to the timestep of the input data. Multiple hours of output data can be averaged to provide estimates across different time periods. WindTrax also outputs a standard deviation in emission rate for each hour of data. The model error within WindTrax is calculated by dividing particles into 10 subgroups and assessing the standard deviation of the mean value estimated using each subgroup. As in the case of the model error output by Airviro, it is unlikely that the WindTrax model error represents a complete uncertainty budget, not least because it doesn't include uncertainty within

620   the measured input parameters.

In general, the model used in WindTrax assumes that there is an equal number of concentration measurements (i.e., sampling locations) and unknown emission sources. If the number of concentration measurements is greater than the number of unknown sources (as is the case in the validation set up here), the simultaneous equations are solved using a least-squares fit.

Custom Gaussian plume model

CH$_4$ emissions were also quantified using a simple custom-built Gaussian plume analysis, similar in concept to that described in OTM-33A (US EPA, 2014). OTM-33A is applicable for ground-level emission sources located within 200 m of the receptor. OTM-33A has provision for stationary measurements made within an emission plume, and this has been utilised in several emission studies (Foster-Wittig et al., 2015; Shaw et al., 2020). Variations in wind direction move the plume around the stationary location in three dimensions over the sampling period. A point-source Gaussian analysis is then employed to derive the emission rate. For the purposes of this work, the typical approach to emission quantification prescribed in OTM-33A was adapted to the use of multiple stationary receptors. The wind direction was rotated to the frame of reference for each individual sampling inlet location, with a rotated wind direction of zero representative of a wind direction blowing over the controlled release location and directly towards the sampling inlet location. CH$_4$ concentration data were then binned into 2° bins (by rotated wind direction) and averaged. A Gaussian function was fit to the averaged concentration data.

The maximum CH$_4$ concentration output by the Gaussian function ($a1$) was used to estimate the CH$_4$ emission rate, using Eq. B1 (US EPA, 2014).

$$F = a1 \times \iint \exp - \left(\frac{y^2}{2\sigma_y{}^2} + \frac{z^2}{2\sigma_z{}^2} \right) = 2\pi \times a1 \times \sigma_y \times \sigma_z \times u \qquad \text{Eq. B1}$$

Where F is the CH$_4$ emission flux, $a1$ is the maximum CH$_4$ concentration output from the Gaussian function, u is the mean wind speed, and σ_y and σ_z are the standard deviations of the crosswind (y) and vertical (z) plume spread respectively. Included within OTM-33A is provision for determining the Pasquill atmospheric stability class during measurements using the measured standard deviation in wind direction (σ_{wd}). The stability class, along with the distance between the release point and the receptor, are in turn used to determine values for σ_y and σ_z from a look-up table (US EPA, 2014). Using one-minute measured values of σ_{wd}, the atmospheric stability class was calculated to be '*Extremely stable*' (Class 7) for 82%, 91%, 92%, and 94% of the duration of R1, R2, R3, and R4 respectively (see Appendix A).

Uncertainties in the emission rate were calculated by propagation of the uncertainties in each element in Eq. B1 in quadrature. When assessing the uncertainty in σ_y and σ_z, the difference between the σ_y and σ_z values for Class 7 and the adjacent stability class (Class 6, '*Moderately stable*') was used as a proxy for uncertainty in σ_y and σ_z.

A high temporal density of data is typically required to generate the required Gaussian function with this method. For this reason, minute-averaged CH$_4$ concentration data were used with the Gaussian plume model (and not hourly-averaged data as was used with both Airviro and WindTrax).

28

Appendix C: Additional results

Figure C1 shows time series of emission rates estimated by Airviro and WindTrax for all tested timesteps and averaging periods, compared with the controlled releases.
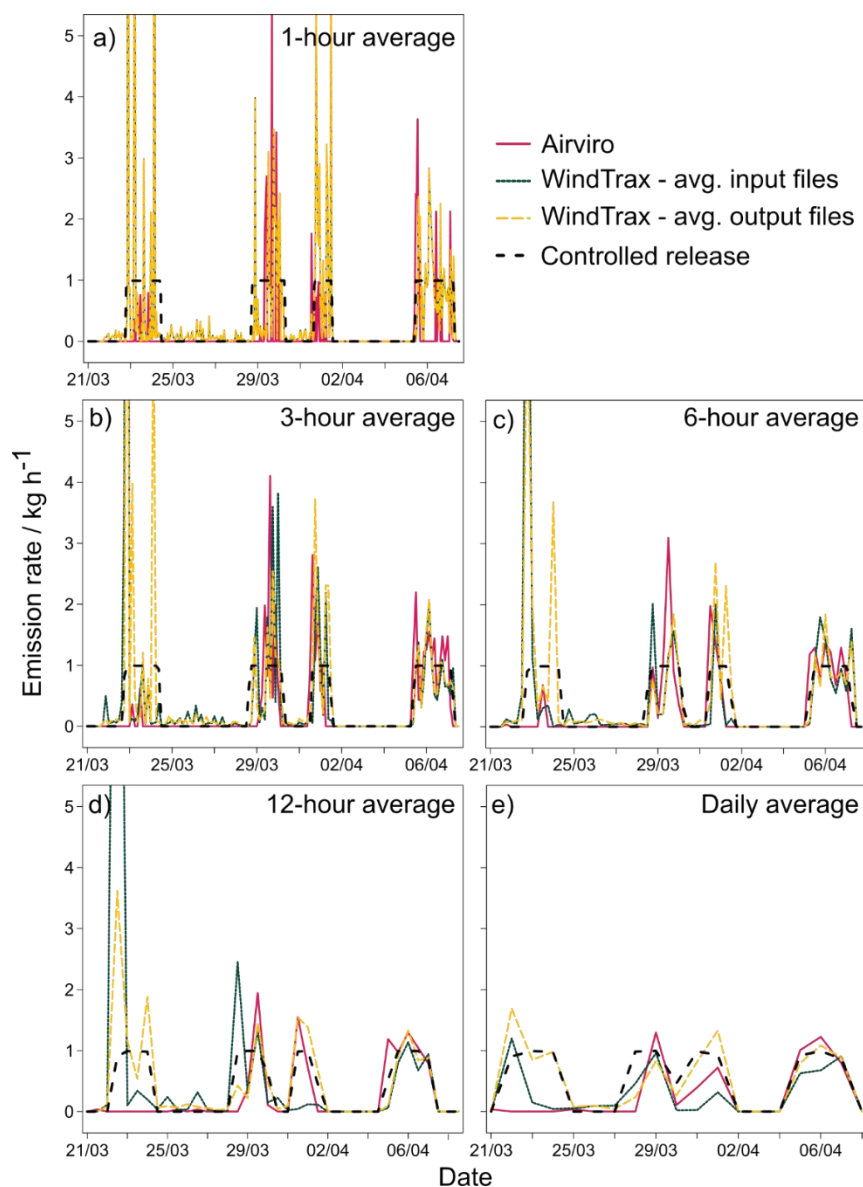

Figure C1. Time series of time averaged emission rate results for Airviro and WindTrax (averaging either the input data or 1-hourly output results). The black dashed line shows time averaged controlled release rates. Time averaging periods used for a) hourly-average, b) 3-hour average, c) 6-hour average, d) 12-hour average, and e) daily-average. Note that the results for averaging the input and output files for hourly-averaging for WindTrax are identical.

660

665

Figure C2 shows the impact of wind persistence on the quantified emission rates.
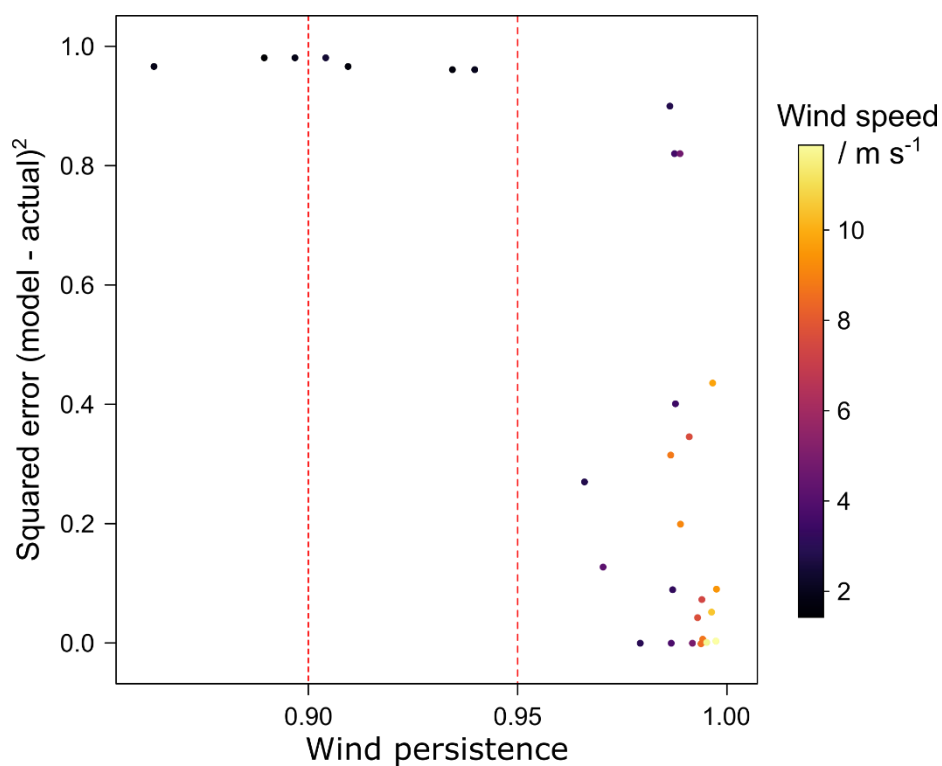
Figure C2. The impact of wind persistence on Airviro estimated emissions, where Airviro results are represented by the square of the error between the modelled emission rate and the controlled emission rate. Data points are shown for daily and 12-hourly time averages. Data points coloured by wind speed (in m s⁻¹). Two wind persistence thresholds are shown by the dashed red lines, equivalent to wind persistence values of 0.95 and 0.90.

670

675 Table C1 shows the same results as those in Table 4, but with periods where average hourly wind persistence was below a threshold of 0.95 (<0.95) removed. Removing periods of lower wind persistence improved the success index for Airviro, particularly for longer averaging periods, with daily and 12-hourly time steps over 90% successful at detecting the occurrence (or non-occurrence) of emissions. The Airviro bias also improved after removing periods of low wind persistence, with all averaging periods (except hourly) showing a mean bias closer to zero. The RMSE decreased somewhat for daily and 12-hourly

680 averaging periods but increased marginally for 6-hourly and 3-hourly averaging. The only timestep which showed little-to-no improvement over including periods of low wind persistence for Airviro was the hourly averaging. This was due to the large proportion of no emissions calculated even when wind persistence was reasonably high (>0.95).

The picture for WindTrax was slightly different after removing periods of low wind persistence. The model performance actually marginally decreased in some cases. This was largely because WindTrax was able to output an emission

685 rate even during periods of low wind persistence, where Airviro struggled.

Table C1. Metrics comparing the Airviro and WindTrax modelled emissions rates with the controlled release emission rates whilst varying the time step (or averaging period). Periods of lower wind persistence (<0.95) were removed from the data relative to Table 4 (see Fig. C3 for contingency matrix pre- and post-application of the wind persistence threshold). Values in brackets are for a wind

690 **persistence threshold of <0.90. See Figure 8 for graphical summary.**

Model	Time step (or averaging period)	Success index / %	Bias / kg h^{-1}	RMSE / kg h^{-1}
Airviro	Daily	94 (85)	-0.13 (-0.26)	0.37 (0.51)
	12-hourly	92 (86)	-0.01 (-0.09)	0.41 (0.47)
	6-hourly	85 (83)	-0.06 (-0.08)	0.54 (0.53)
	3-hourly	80 (79)	-0.09 (-0.10)	0.65 (0.64)
	Hourly	60 (60)	-0.31 (-0.32)	0.77 (0.76)
WindTrax	Daily	67 (67)	0.06 (-0.01)	0.29 (0.34)
	12-hourly	61 (64)	0.17 (0.16)	0.66 (0.64)
	6-hourly	57 (59)	0.18 (0.22)	1.05 (1.08)
	3-hourly	63 (64)	0.17 (0.15)	1.32 (1.27)
	Hourly	68 (67)	0.17 (0.15)	2.19 (2.14)

Figure C3 shows contingency matrices for both Airviro and WindTrax when using no wind persistence threshold (all data) and when using a wind persistence threshold value of 0.95. The percentage of true positives increased for Airviro (with the threshold applied) in all cases with different time averaging except for hourly. For Airviro, this was largely a result of the removal of data during R1 when the wind persistence was generally below 0.95. The impact of using a wind persistence threshold on the contingency metrics for WindTrax was much less pronounced, as WindTrax output emission rates during R1.
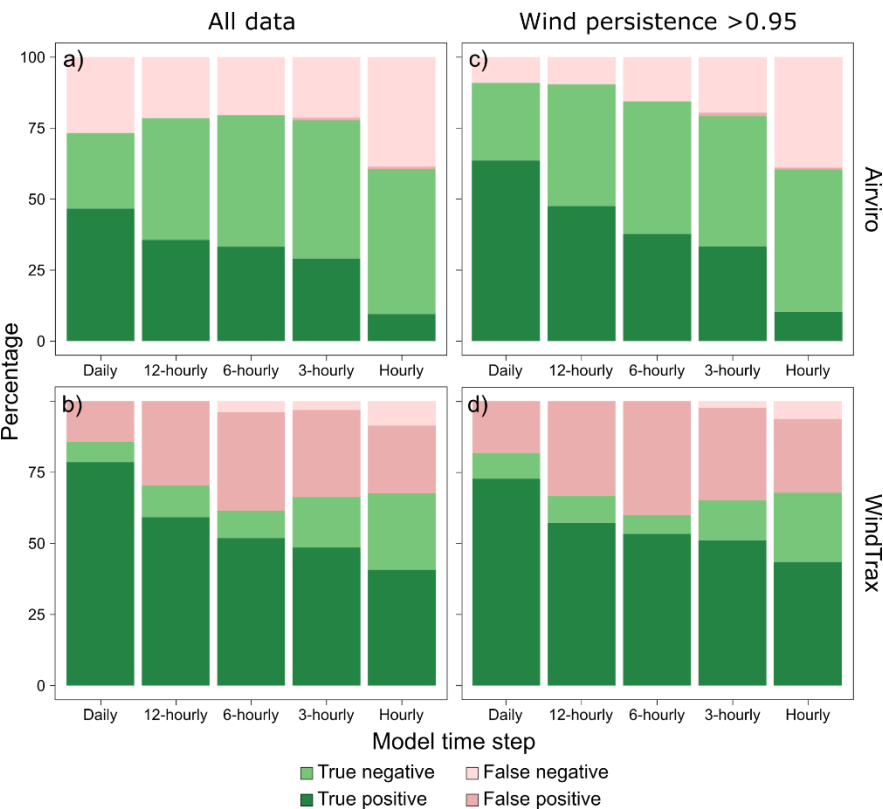

Figure C3. Contingency matrices (percentages of true positives, true negatives, false positives, and false negatives) for Airviro (top panels) and WindTrax (bottom panels) when using all data (left panels) and when using a wind persistence threshold of 0.95 (right panels). All results shown with a lag time of 60 seconds and a plume height of 2 m.

32

Table C2. Average values for wind persistence, controlled release rate (kg h⁻¹), and Airviro and WindTrax estimated emission rates (with and without a wind persistence threshold of 0.95 applied) over the periods covering each of the four controlled releases and under different time averaging scenarios. Note that the specific start and end dates for releases impacted the average CRF rate over different time periods; if, for a non-realistic example, the release covered the period 12:00-23:59, the average daily emission rate covering that period would be 0.5 kg h⁻¹ whereas the average 12-hourly rate for that period would be 1 kg h⁻¹.

710

Time	Release number	R1	R2	R3	R4
Daily	Wind persistence	0.94		0.97ᵃ	1.00
	CRF emission rate (kg h⁻¹)	0.96 ± 0.05		0.87 ± 0.23^a	0.92 ± 0.07
	Airviro (kg h⁻¹)	0		0.50 ± 0.52^a	1.01 ± 0.22
	Airviro w/ wind pers. threshold (kg h⁻¹)	0		0.63 ± 0.51^a	1.01 ± 0.22
	WindTrax (kg h⁻¹)	1.17 ± 0.45		0.70 ± 0.46^a	0.85 ± 0.21
	WindTrax w/ wind pers. threshold (kg h⁻¹)	1.69		0.81 ± 0.43^a	0.85 ± 0.21
12-hourly	Wind persistence	0.93	0.95	0.99	1.00
	CRF rate	0.97 ± 0.04	0.90 ± 0.18	0.81 ± 0.31	0.87 ± 0.20
	Airviro	0	0.60 ± 0.91	0.76 ± 0.77	1.06 ± 0.19
	Airviro w/ wind pers. threshold	0	0.80 ± 1.00	0.77 ± 0.76	1.06 ± 0.19
	WindTrax	1.80 ± 1.33	0.64 ± 0.54	1.20 ± 0.47	0.77 ± 0.45
	WindTrax w/ wind pers. threshold	3.62	0.72 ± 0.64	1.20 ± 0.47	0.77 ± 0.45
6-hourly	Wind persistence	0.92	0.95	0.99	1.00
	CRF rate	0.98 ± 0.03	0.86 ± 0.27	0.88 ± 0.24	0.88 ± 0.22
	Airviro	0.10 ± 0.26	0.81 ± 1.02	0.84 ± 0.86	0.96 ± 0.44
	Airviro w/ wind pers. threshold	0	1.09 ± 1.05	0.84 ± 0.86	0.96 ± 0.44
	WindTrax	2.04 ± 2.59	0.63 ± 0.64	1.30 ± 1.11	0.84 ± 0.47
	WindTrax w/ wind pers. threshold	2.41 ± 3.26	0.84 ± 0.60	1.30 ± 1.11	0.84 ± 0.47
3-hourly	Wind persistence	0.92	0.96	0.99	1.00
	CRF rate	0.97 ± 0.05	0.94 ± 0.15	0.92 ± 0.19	0.93 ± 0.17
	Airviro	0.07 ± 0.18	0.77 ± 1.18	0.90 ± 1.00	1.03 ± 0.58
	Airviro w/ wind pers. threshold	0.04 ± 0.12	0.83 ± 1.21	0.90 ± 1.00	1.03 ± 0.58
	WindTrax	1.88 ± 3.50	0.72 ± 0.76	1.52 ± 1.21	0.90 ± 0.49
	WindTrax w/ wind pers. threshold	2.00 ± 3.85	0.78 ± 0.76	1.52 ± 1.21	0.90 ± 0.49
Hourly	Wind persistence	0.92	0.98	0.99	1.00
	CRF rate	0.97 ± 0.14	0.97 ± 0.09	0.97 ± 0.07	0.95 ± 0.17
	Airviro	0.05 ± 0.18	0.47 ± 1.25	0.17 ± 0.33	0.40 ± 0.83
	Airviro w/ wind pers. threshold	0.03 ± 0.16	0.50 ± 1.28	0.18 ± 0.34	0.40 ± 0.83

WindTrax	1.96 ± 6.20	0.78 ± 1.03	1.67 ± 1.86	0.94 ± 0.70
WindTrax w/ wind pers. treshold	1.96 ± 6.65	0.82 ± 1.05	1.78 ± 1.92	0.94 ± 0.70

[a] Release 2 (R2) ended on the same day that Release 3 (R3) began. Thus, if using daily averaging (between 00:00 and 23:59), the two releases could not be separated in time. This was not a problem for shorter averaging periods (e.g., 12-hourly) as R2 finished in the morning (before 12:00) and R3 began in the afternoon (after 12:00).

715

34

The lag time refers to the period of time over which data was excluded following switching between sampling inlet locations. There was unlikely to be instantaneous measurement of air sampled from the new location due to a lag in the flushing time of the optical cell. However, the longer the chosen lag time, the more data invalidated and the smaller proportion of quality data available for analysis and model interpretation.

720   Airviro and WindTrax were tested using different lag periods to understand if this had any impact on quantified emission rates. Varying the lag time seemed to have only a minor effect on the success index for both models (see Table C3). Figure C4 shows the bias and RMSE scores achieved whilst varying the sampling lag time for both Airviro and WindTrax. Shorter lag times yielded smaller biases for Airviro but larger values for RMSE, indicating that an intermediate value for lag time may be optimal for Airviro. The picture for WindTrax was more mixed with no clear trend. The lowest bias and RMSE

725 values for WindTrax were observed when using a lag time of 120 s, with higher bias and RMSE scores observed when using a lag time of 75 s. Overall, varying the length of the lag period between sampling inlets had only a minor impact on the magnitude of emission rates estimated by both models.
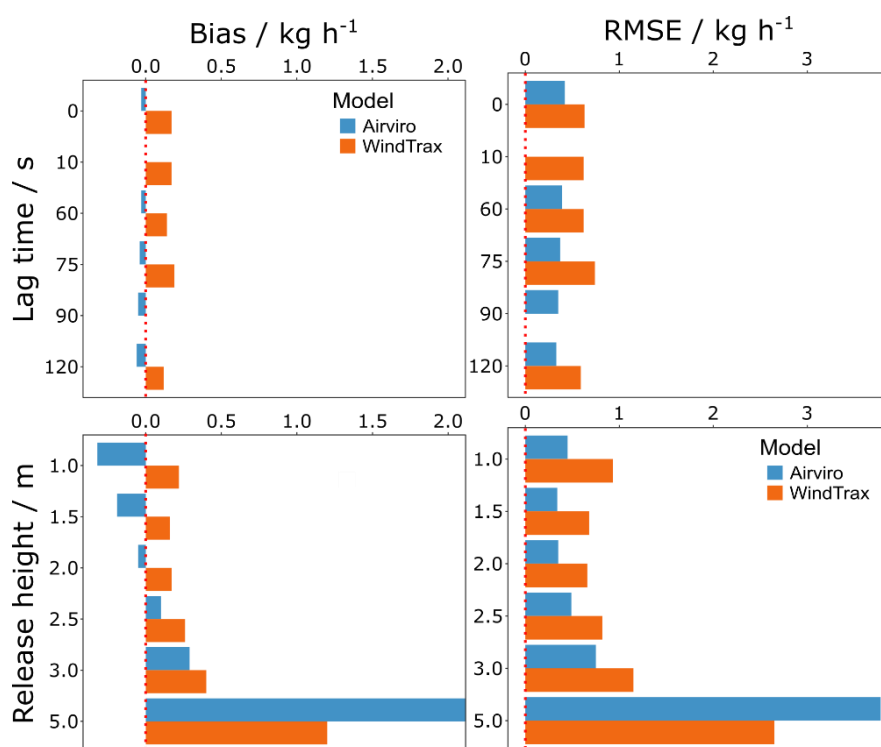
730 **Figure C4. Graphical summary of bias (left column) and RMSE (right column) for Airviro (blue) and WindTrax (orange) under model simulations with different lag times (top row) and release heights (bottom row). The lag time refers to the length of time that CH$_4$ concentration data was invalidated following switching of sampling inlets (to preclude the possibility of measuring air sampled from the previous sampling inlet). The release height (or plume height) refers to the height of the simulated release point. Ideal values are indicated by red dashed lines (success index = 100%; bias = 0 kg h^{-1}; RMSE = 0 kg h^{-1}). See Table C3 for values.**

735   Both Airviro and WindTrax require the user to specify the height at which gas was emitted. The CRF released CH_4 from a

height of 2 m above ground level. However, knowledge of the release height may not be known in a real-world scenario,

particularly if the emission source is unknown or if there are multiple emission sources in proximity. Figure C4 shows the

sensitivity of both models to the release height (when using 12-hourly time periods and a wind persistence threshold of 0.95;

all other variables constant). For context, the true release height of the emission plume was 2.0 m. See Table C3 for success

740   index results.

Both models performed better overall when the release height was set to the actual value of 2.0 m. Airviro and

WindTrax both yielded lower bias and RMSE values for simulations at 2.0 m, although some bias and RMSE scores were

marginally lower for simulations with plume heights within 0.5 m of the true release height. The bias and RMSE values for

both models increased as the simulated plume height was set at greater distances from the true release height, with the worst

745   performances when the simulated plume was set at 5.0 m. The results therefore imply that small discrepancies in emission

plume height (±0.5 m in this case) may not substantially alter the final results, but that discrepancies greater than this value

(e.g., ≥±1.0 m) can result in substantial model errors. More controlled releases tests at different heights would need to be

undertaken to evaluate the influence of plume height in greater detail.

Table C3 shows the success index, bias, and RMSE results for Airviro and WindTrax whilst varying the lag time (in

750   seconds) and the release height (in m). See Fig. C4 for a graphical representation of these results. Varying the lag time (the

length of time over which CH_4 concentration data was invalidated following switching of sampling inlets) had only a marginal

impact on the success index, bias, and RMSE results for both models. A lag time of between 60 and 90 seconds appeared to

be optimal. The model setting for release height had a much more substantial impact on the quantified emissions. The results

for Airviro showed the model switched bias, with Airviro underestimating emissions for release heights at or below the true

755   release height, but overestimating emissions for release heights greater than the true release height.

Table C3. Metrics comparing the Airviro and WindTrax modelled emissions rates with the controlled release emission rates whilst varying different parameters. The lag time refers to the length of time that CH$_4$ concentration data was invalidated following switching of sampling inlets (to preclude the possibility of measuring air sampled from the previous sampling inlet). The release height (or plume height) refers to the height of the release point.

Model	Lag time / s	Success index / %	Bias / kg h^{-1}	RMSE / kg h^{-1}
Airviro	0	88	-0.03	0.42
	60	92	-0.03	0.39
	75	92	-0.04	0.37
	90	92	-0.05	0.35
	120	92	-0.06	0.33
WindTrax	0	64	0.17	0.63
	10	64	0.17	0.62
	60	61	0.14	0.62
	75	64	0.19	0.74
	120	64	0.12	0.59
Model	**Release height / m**	**Success index / %**	**Bias / kg h^{-1}**	**RMSE / kg h^{-1}**
Airviro	1.0	92	-0.32	0.45
	1.5	92	-0.19	0.34
	2.0 [a]	92	-0.05	0.35
	2.5	88	0.10	0.49
	3.0	88	0.29	0.75
	5.0	71	2.21	3.78
WindTrax	1.0	61	0.22	0.93
	1.5	61	0.16	0.68
	2.0*	61	0.17	0.66
	2.5	61	0.26	0.82
	3.0	56	0.40	1.15
	5.0	50	1.20	2.65

[a] True release height.

Table C4 shows the results when varying the number of active sampling inlets used for input into Airviro and WindTrax simulations. Figure C5 also shows schematic configurations of the chosen sampling inlet locations around the CRF release point location for each model scenario.

765

Table C4. Metrics comparing the Airviro and WindTrax modelled emissions rates with the controlled release emission rates whilst varying the number of active inlets used for input of CH$_4$ concentration data. See Fig. C5 for schematic configurations of inlet locations.

Scenario	Active sampling inlets	Success index / %		RMSE / kg h^{-1}	
		Airviro	WindTrax	Airviro	WindTrax
12a	All 12	92	61	0.35	0.64
11a	Excluding no. 1	83		0.50	
11b	Excluding no. 12	88		0.40	
11c	Excluding no. 6	92		0.36	
11d	Excluding no. 2	92		0.38	
11e	Excluding no. 9	92		0.36	
11f	Excluding no. 8	92		0.37	
11g	Excluding no. 5	92		0.39	
9a	Excluding nos. 1, 7, 11	79	61	0.56	1.37
9b	Excluding nos. 2, 6, 12	88	67	0.41	0.58
9c	Excluding nos. 3, 8, 9	50		0.65	
9d	Excluding nos. 4, 5, 10	50		0.67	
8a	Excluding nos. 3, 4, 11, 12	88	61	0.56	1.51
8b	Excluding nos. 1, 5, 6, 9	83	56	0.53	1.59
8c	Excluding nos. 2, 7, 8, 10	88	56	0.51	0.85
6a	Nos. 1, 3, 7, 8, 9, 11	88	67	0.68	0.82
6b	Nos. 2, 4, 5, 6, 10, 12	79	61	0.61	1.73
3a	Nos. 1, 7, 11	75		0.92	
3b	Nos. 4, 5, 10	58	43	0.65	0.46
3c	Nos. 3, 8, 9	58	64	0.75	0.43
3d	Nos. 2, 6, 12	71		0.56	
1a	No. 1	57		2.37	
1b	No. 12	44		0.67	
1c	No. 6	51		0.59	
1d	No. 2	43		0.67	

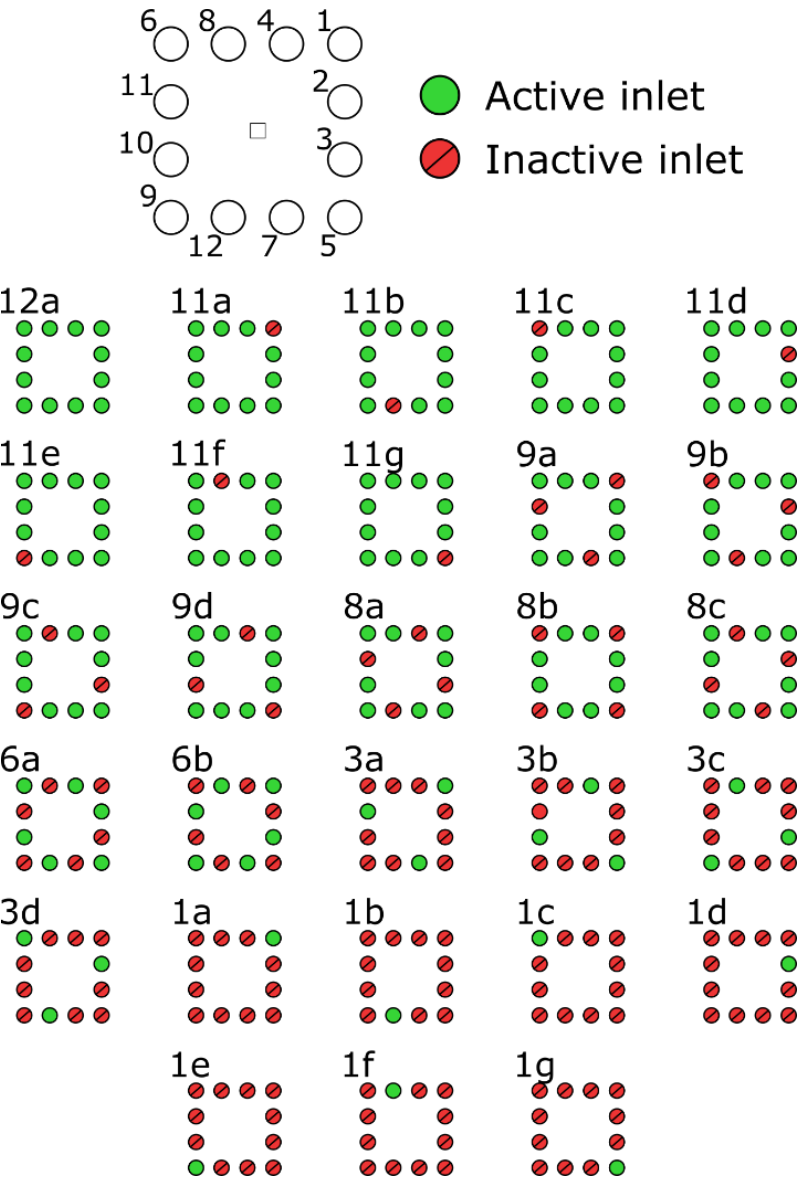| 1e | No. 9 | 53 | 0.64 |
| 1f | No. 8 | 62 | 0.63 |
| 1g | No. 5 | 49 | 0.65 |

Figure C5. Schematics of scenarios in which differing configurations of sampling inlet locations were input for modelling scenarios (see Table C4). See also Fig. 2.

WindTrax requires the user to specify a value for the local CH_4 background concentration. The CH_4 background concentration

775 can be a difficult value to prescribe accurately, especially in the presence of extraneous emission sources and also in the context of long-term monitoring over which the background might be expected to vary with season (e.g., Shaw et al., 2019). Airviro has no user-specified setting for the CH_4 background and instead calculates its own background value internally.

Table C5 shows the results of varying the CH_4 background setting for WindTrax between 0 ppm and 2 ppm, with 2 ppm representative of a typical UK background at the time of publication. A lower CH_4 background value resulted in WindTrax

780 overestimating emissions, as might be expected. WindTrax performed best when the CH_4 background was set to a value which could be considered reasonably representative of the true background (the 5th percentile of CH_4 mole fractions measured during the campaign was 2.002 ppm). More simulations varying the background value across a range of actually feasible background methane concentrations (e.g., 1.90 – 2.10 ppm) would be instructive for determining the sensitivity of WindTrax to the choice of background value.

785

Table C5. Metrics comparing WindTrax modelled emissions rates with the controlled release emission rates whilst varying the CH_4 background concentration.

| Model	CH_4 background concentration / ppm	Success index / %	Bias / kg h^{-1}	RMSE / kg h^{-1}
WindTrax	0	50	3.85	4.88
	1	50	2.00	2.51
	2	57	0.05	0.55

790 Airviro does not have capability to simulate single point emission sources and instead requires the user to designate an emission source zone from which emission rates are then calculated. WindTrax, on the other hand, allows for the user to specify the location of known point sources. Figure C6 shows a number of different emission zones which were tested with Airviro to assess the sensitivity of the model to the size of emission zone. It should be noted that the default emission zone used for all other scenarios was Emission Zone 3.

795 The success index decreased with increasing size of the emission zone, especially beyond the circle of sampling locations. When the emission zone was kept to mostly within the circle of sampling locations, the bias was negative. However, the bias was positive if the emission zone was extrapolated outside of the circle of sampling inlets. RMSE was lower for emission zones within the circle, and greater for emission zones that extended beyond the circle. This implies that Airviro was best able to capture emissions when multiple sampling locations surrounded the emission source.

800

Table C6. Metrics comparing Airviro modelled emission rates with the controlled release emission rates whilst varying the emission zone (see Fig. C6).

Model	Emission zone	Success index / %	Bias / kg h⁻¹	RMSE / kg h⁻¹
Airviro	1	92	-0.20	0.39
	2	92	-0.15	0.33
	3	92	-0.05	0.35
	4	65	0.58	1.17
	5	50	1.05	2.07
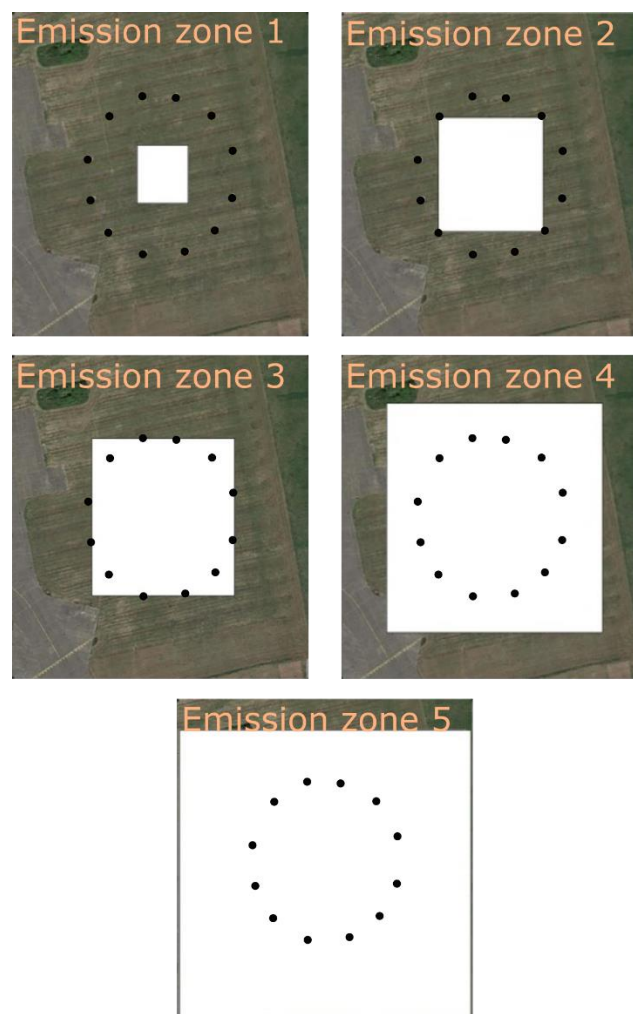
805 **Figure C6. Configuration of five different emission zones (white areas) tested with Airviro. The black dots indicate the locations of the sampling inlets (see Fig. 2). Emission zone 3 was used for all other Airviro simulations in this work. Satellite imagery: Google, ©2020 Maxar Technologies.**

WindTrax allows the user to specify the number of particles modelled for dispersion for each time step. Increasing the number
810 of particles theoretically increases the accuracy of dispersion modelling but this comes at the expense of increased
computational power and calculation time.

Table C7 shows the results of varying the particle number for WindTrax between 5,000 particles and 100,000
particles. The simulation with the fewest particles (5,000) resulted in a negative bias whereas all other simulations yielded
positive biases. The simulation with 5,000 particles also resulted in the lowest value for RMSE. The simulation with 10,000
815 particles gave the greatest success index and lowest bias but one of the higher values for RMSE. Increasing the particle number
beyond 30,000 particles led to no improvement in success index and only marginal improvements in bias. Increasing the
particle number beyond 20,000 led to a general decrease in RMSE up to 50,000 particles. 50,000 particles were used in most
other WindTrax simulations in this report as it was seen as a compromise between reasonable metric scores and short
calculation time.

820

Table C7. Metrics comparing WindTrax modelled emissions rates with the controlled release emission rates whilst varying the particle number.

Model	Particle number	Success index / %	Bias / kg h^{-1}	RMSE / kg h^{-1}
WindTrax	5,000	75	-0.34	0.52
	10,000	79	0.07	0.89
	20,000	57	0.22	0.97
	30,000	61	0.21	0.81
	40,000	61	0.19	0.71
	50,000	61	0.17	0.66
	100,000	61	0.19	0.70

Data availability

Data may be made available upon request.

Author contributions

Author contributions using CRediT (Contributor Roles Taxonomy).

JTS: Formal analysis, Methodology, Validation, Visualization, Writing – original draft preparation.

NH: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft preparation.

JC: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation.

DEB: Formal analysis, Visualization, Writing – original draft preparation.

JR: Data curation, Investigation, Software.

JH: Investigation, Validation, Writing – original draft preparation.

NY: Investigation, Validation, Formal analysis.

DB: Conceptualization, Methodology, Supervision.

FI: Conceptualization, Methodology, Supervision.

RR: Conceptualization, Methodology, Project administration, Supervision, Writing – original draft preparation.

Competing interests

The FEDS system is a commercial property owned and operated by NPL.

Acknowledgements

Financial support

References

Allen, G.: How to quantify anthropogenic methane emissions with aircraft surveys, Commun. Earth Environ., 4, 140, https://doi.org/10.1038/s43247-023-00794-6, 2023.

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D. and Andreassian, V.: Characterising performance of environmental models, Env. Model. Soft., 40, 1-20, https://doi.org/10.1016/j.envsoft.2012.09.011, 2013.

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, https://doi.org/10.1016/j.atmosenv.2008.06.019, 2008.

Connor, A., Shaw, J. T., Yarrow, N., Howes, N., Helmore, J., Finlayson, A., Barker, P. and Robinson, R.: A framework for describing and classifying methane reporting requirements, emission sources, and monitoring methods, Env. Sci. Atmos., https://doi.org/10.1039/D4EA00120F, 2024.

Cusworth, D. H., Duren, R. M., Yadav, V., Thorpe, A. K., Verhulst, K., Sander, S., Hopkins, F., Rafiq, T. and Miller, C. E.: Synthesis of methane observations across scales: Strategies for deploying a multitiered observing network, Geophys. Res. Lett., 47, e2020GL087869, https://doi.org/10.1029/2020GL087869, 2020.

Defratyka, S. M., France, J. L., Fisher, R. E., Lowry, D., Fernandez, J. M., Bakkaloglu, S., Yver-Kwok, C., Paris, J.-D., Bousquet, P., Arnold, T., Rennick, C., Helmore, J., Yarrow, N. and Nisbet, E. G.: Statistical evaluation of methane isotopic signatures determined during near-source measurements, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-1490, 2023.

EN 17628:2022, Fugitive and diffuse emissions of common concern to industry sectors – Standard method to determine diffuse emissions of volatile organic compounds into the atmosphere, CEN/TC 264 – Air Quality, WG 38 – Determination of fugitive VOC emissions, 2022.

Erland, B. M., Thorpe, A. K. and Gamon, J. A.: Recent advances toward transparent methane emissions monitoring: A review, Environ. Sci. Technol., 56, 23, 16567-16581, https://doi.org/10.1021/acs.est.2c02136, 2022.

European Commission, Launch by United States, the European Union, and Partners of the Global Methane Pledge to Keep 1.5C Within Reach, Statement/21/5766/, https://ec.europa.eu/commission/presscorner/detail/en/statement_21_5766, Last accessed: May 2023, 2021.

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, 1320-1332, https://doi.org/10.1175/1520-0450(1995)034%3C1320:BTLSDM%3E2.0.CO;2, 1995.

885   Flesch, T. K., Harper, L. A., Powell, J. M. and Wilson, J. D.: Inverse-dispersion calculation of ammonia emissions from Wisconsin dairy farms, Trans. ASABE, 52, 253-265, https://doi.org/10.13031/2013.25946, 2009.

Foster-Wittig, T. A., Thoma, E. D. and Alberston, J. D.: Estimation of point source fugitive emission rates from a single sensor time series: A conditionally-sampled Gaussian plume reconstruction, Atmos. Environ., 115, 101-109, https://doi.org/10.1016/j.atmosenv.2015.05.042, 2015.

890   Fox, T. A., Barchyn, T. E., Risk, D., Ravikumar, A. P. and Hugenholtz, C. H.: A review of close-range and screening technologies for mitigating fugitive methane emissions in upstream oil and gas, Environ. Res. Lett., 14, 5, 053002, https://doi.org/10.1088/1748-9326/ab0cc3, 2019.

Gardiner, T., Helmore, J., Innocenti, F. and Robinson, R.: Field validation of remote sensing methane emission measurements, Remote Sens., 9, 9, 956, https://doi.org/10.3390/rs9090956, 2017.

895   Innocenti, F., Robinson, R., Gardiner, T., Howes, N. and Yarrow, N.: Comparative assessment of methane emissions from onshore LNG facilities measured using differential absorption LiDAR, Environ. Sci. Technol., 57, 3301-3310, https://doi.org/10.1021/acs.est.2c05446, 2023.

Jacob, D. J., Varon, D. J., Cusworth, D. H., Dennison, P. E., Frankenberg, C., Gautam, R., Guanter, L., Kelley, J., McKeever, J., Ott, L. E., Poulter, B., Qu, Z., Thorpe, A. J., Worden, J. R. and Duren, R. M.: Quantifying methane emissions from 900   the global scale down to point sources using satellite observations of atmospheric methane, Atmos. Chem. Phys., 22, 14, 9617-9646, https://doi.org/10.5194/acp-22-9617-2022, 2022.

Jones, A. R., Thomson, D. J., Hort, M. and Devinish, B.: The UK Met Office's next generation atmospheric dispersion model, NAME III, in Borrego, C. and Norman, A.-L. (Eds) *Air Pollution Modeling and its Application XVII*, Proceedings of the 27th NATO/CCMS International Technical Meeting on Air Pollution Modelling and its Application, Spring, 580-905   589, 2007.

Kumar, P., Broquet, G., Caldow, C., Laurent, O., Gichuki, S., Cropley, F., Yver-Kwok, C., Fontanier, B., Lauvaux, T., Ramonet, M., Shah, A., Berthe, G., Martin, F., Duclaux, O., Juery, C., Bouchet, C., Pitt, J., Ciais, P.: Near-field atmospheric inversions for the localization and quantification of controlled methane releases using stationary and mobile measurements, Q. J. Roy. Meteor. Soc., 148, 1886-1912, https://doi.org/10.1002/qj.4283, 2022.

910   Nisbet, E. G., Fisher, R. E., Lowry, D., France, J. L., Allen, G., Bakkaloglu, S., Broderick, T. J., Cain, M., Coleman, M., Fernandez, J., Forster, G., Griffiths, P. T., Iverach, C. P., Kelly, B. F. J., Manning, M. R., Nisbet-Jones, P. B. R., Pyle, J. A., Townsend-Small, A., al-Shalaan, A., Warwick, N. and Zazzeri, G.: Methane mitigation: Methods to reduce emissions, on the Path to the Paris Agreement, Rev. Geophys., 58, 1, e2019RG000675, https://doi.org/10.1029/2019RG000675, 2020.

915   Nisbet, E. G.: The urgent need to cut methane emissions, Natl. Sci. Rev., 9, 5, https://doi.org/10.1093/nsr/nwab221, 2022a.

Nisbet, E. G.: Methane's unknowns better known, Nat. Geosci., 15, 861-862, https://doi.org/10.1038/s41561-022-01049-3, 2022b.

Ocko, I. B., Sun, T., Shindell, D., Oppenheimer, M., Hristov, A. N., Pacala, S. W., Mauzerall, D. L., Xu, Y. and Hamburg, S. P.: Acting rapidly to deploy readily available methane mitigation measures by sector can immediately slow global
920   warming, Env. Res. Lett., 16, 054042, https://doi.org/10.1088/1748-9326/abf9c8, 2021.

Olczak, M., Piebalgs, A. and Balcombe, P.: A global review of methane policies reveals that only 13% of emissions are covered with unclear effectiveness, One Earth, 6, 5, 519-535, https://doi.org/10.1016/j.oneear.2023.04.009, 2023.

Pisso, I., Sollum, E., Grythe, H., Kristiansen, N. I., Cassiani, M., Eckhardt, S., Arnold, D., Morton, D., Thompson, R. L., Zwaaftink, C. D. G., Evangeliou, N., Sodemann, H., Haimberger, L., Henne, S., Brunner, D., Burkhart, J. F.,
925   Fouilloux, A., Brioude, J., Philipp, A., Seibert, P. and Stohl, A.: The Lagrangian particle dispersion model FLEXPART version 10.4, Geosci. Model Dev., 12, 4955-4997, https://doi.org/10.5194/gmd-12-4955-2019, 2019.

Riddick, S. N., Connors, S., Robinson, A. D., Manning, A. J., Jones, P. S. D., Lowry, D., Nisbet, E., Skelton, R. L., Allen, G., Pitt, J. and Harris, N. R. P.: Estimating the size of a methane emission point source at different scales: from local to landscape, Atmos. Chem. Phys., 17, 12, 7839-7851, https://doi.org/10.5194/acp-17-7839-2017, 2017.

930   Riddick, S. N., Mauzerall, D. L., Celia, M., Allen, G., Pitt, J., Kang, M. and Riddick, J. C.: The calibration and deployment of a low-cost methane sensor, Atmos. Environ., 230, 117440, https://doi.org/10.1016/j.atmosenv.2020.117440, 2020.

Riddick, S. N., Ancona, R., Cheptonui, F., Bell, C. S., Duggan, A., Bennett, K. E. and Zimmerle, D. J.: A cautionary report of calculating methane emissions using low-cost fence-line sensors, Elem. Sci. Anth., 10, 1, 00021, https://doi.org/10.1525/elementa.2022.00021, 2022.

935   Rivera-Martinez, R., Kumar, P., Laurent, O., Broquet, G., Caldow, C., Cropley, F., Santaren, D., Shah, A., Mallet, C., Ramonet, M., Rivier, L., Juery, C., Duclaux, O., Bouchet, C., Allegrini, E., Utard, H. and Ciais, P.: Using metal oxide gas sensors to estimate the emission rates and locations of methane leaks in an industrial site: Assessment with controlled methane releases, Atmos. Meas. Tech., 17, 4257-4290, https://doi.org/10.5194/amt-17-4257-2024, 2024.

Shah, A., Allen, G., Pitt, J. R., Ricketts, H., Williams, P. I., Helmore, J., Finlayson, A., Robinson, R., Kabbabe, K.,
940   Hollingsworth, P., Rees-White, T. C., Beaven, R., Scheutz, C. and Bourn, M.: A near-field Gaussian plume inversion flux quantification method, applied to unmanned aerial vehicle sampling, Atmosphere, 10, 7, 396, https://doi.org/10.3390/atmos10070396, 2019.

Shaw, J. T., Allen, G., Pitt, J., Mead, M. I., Purvis, R. M., Dunmore, R., Wilde, S., Shah, A., Barker, P., Bateson, P., Bacak, A., Lewis, A. C., Lowry, D., Fisher, R., Lanoisellé, M. and Ward, R. S.: A baseline of atmospheric greenhouse gases
945   for prospective UK shale gas sites, Sci. Tot. Env., 684, 1-13, https://doi.org/10.1016/j.scitotenv.2019.05.266, 2019.

Shaw, J. T., Allen, G., Pitt, J., Shah, A., Wilde, S., Stamford, L., Fan, Z., Ricketts, H., Williams, P. I., Bateson, P., Barker, P., Purvis, R., Lowry, D., Fisher, R., France, J., Coleman, M., Lewis, A. C., Risk, D. A. and Ward, R. S.: Methane flux from flowback operations at a shale gas site, J. Air Waste Ma., 70, 12, https://doi.org/10.1080/10962247.2020.1811800, 2020.

Shaw, J. T., Shah, A., Yong, H. and Allen, G.: Methods for quantifying methane emissions using unmanned aerial vehicles: a review, Phil. Trans. R. Soc. A, 379, 2210, 20200450, https://doi.org/10.1098/rsta.2020.0450, 2021.

Shirvaikar, V. V.: Persistence of wind direction, Atmos. Environ., 6, 12, 889-898, https://doi.org/10.1016/0004-6981(72)90097-2, 1972.

Tollefson, J.: Scientists raise alarm over 'dangerously fast' growth in atmospheric methane, Nature, https://doi.org/10.1038/d41586-022-00312-2, 2022.

US EPA, Draft "Other Test Method" OTM-33 Geospatial Measurement of Air Pollution, Remote Emissions Quantification (GMAP-REQ), https://www.epa.gov/emc, Last accessed June 2023, 2014.

Wang, J. L., Daniels, W. S., Hammerling, D. M., Harrison, M., Burmaster, K., George, F. C. and Ravikumar, A. P.: Multiscale methane measurements at oil and gas facilities reveal necessary frameworks for improved emissions accounting, Environ. Sci. Technol., 56, 20, 14743-14752, https://doi.org/10.1021/acs.est.2c06211, 2022.

Wickström, K., Ivarsson, P. and Silva, G.: Airviro Specification v3.20, Part II: Appendices, Appendix E, Swedish Meteorological and Hydrological Institute, 2010.

Yang, X., Kuru, E., Zhang, X., Zhang, S., Wang, R., Ye, J., Yang, D., Klemeš, J. J. and Wang, B.: Direct measurement of methane emissions from the upstream oil and gas sector: Review of measurement results and technology advances (2018-2022), J. Clean. Prod., 414, 137693, https://doi.org/10.1016/j.jclepro.2023.137693, 2023.