

1 **Title:** Single-blind test of nine methane-sensing satellite systems from three continents

2 **Authors:** Evan D. Sherwin^{1,a,*}, Sahar H. El Abbadi¹, Philippine M. Burdeau¹, Zhan Zhang¹,
3 Zhenlin Chen¹, Jeffrey S. Rutherford^{1,b}, Yuanlei Chen¹, Adam R. Brandt¹

4 **Author Affiliations:**

5 ¹ Department of Energy Science & Engineering, Stanford University, Stanford, California 94305,
6 United States

7 ^a Present affiliation: Lawrence Berkeley National Laboratory, Berkeley, California, 94720,
8 United States

9 ^b Present affiliation: Highwood Emissions Management, Calgary, Alberta T2P 2V1, Canada

10 * Correspondence: evansherwin@lbl.gov

11

12 **Abstract**

13 Satellite-based remote sensing enables detection and mitigation of large point sources of climate-
14 warming methane. These satellites will have the greatest impact if stakeholders have a clear-eyed
15 assessment of their capabilities. We performed a single-blind test of nine methane-sensing
16 satellites from three continents and five countries, including both commercial and government
17 satellites. Over two months, we conducted 82 controlled methane releases during satellite
18 overpasses. Six teams analyzed the resulting data, producing 134 estimates of methane
19 emissions. Of these, 80 (58%) were correctly identified, with 46 true positive detections (34%)
20 and 34 true negative non-detections (25%). There were 41 false negatives, in which teams
21 missed a true emission, and 0 false positives, in which teams incorrectly claimed methane was
22 present. All eight satellites that were given a nonzero emission detected methane at least once,
23 including the first single-blind evaluation of the EnMAP, Gaofen 5, and Ziyuan 1 systems. In
24 percent terms, quantification error across all satellites and teams is similar to aircraft-based
25 methane remote sensing systems, with 55% of mean estimates falling within $\pm 50\%$ of the
26 metered value. Although teams correctly detected emissions as low as 0.03 metric tons of
27 methane per hour, it is unclear whether detection performance in this test is representative of
28 real-world field performance. Full retrieval fields submitted by all teams suggest that in some
29 cases it may be difficult to distinguish true emissions from background artifacts without a known
30 source location. Cloud interference is significant and appears to vary across teams and satellites.
31 This work confirms the basic efficacy of the tested satellite systems in detecting and quantifying
32 methane, providing additional insight into detection limits and informing experimental design for
33 future satellite-focused controlled methane release testing campaigns.

34
35 **Keywords:**

36 Methane, hyperspectral imaging, remote sensing, satellite, single-blind, controlled release

37
38 Satellite-based remote sensing systems continue to find large point-source emissions of climate-
39 warming methane across the globe (Lauvaux et al., 2022; Irakulis-Loitxate et al., 2022a, b;
40 Pandey et al., 2019; Varon et al., 2018, 2019, 2021; Sánchez-García et al., 2022). Such systems
41 empower stakeholders in industry and government to take corrective action, both to mitigate
42 individual sources and to inform estimates of total methane emissions, particularly in oil and
43 natural gas systems, where many of the largest sources have been observed (Lauvaux et al.,
44 2022; Irakulis-Loitxate et al., 2022a; Pandey et al., 2019; Varon et al., 2018; Irakulis-Loitxate et
45 al., 2022b; Varon et al., 2021, 2019; Cusworth et al., 2022; Duren et al., 2019; Chen et al., 2022;
46 Sherwin et al., 2023a; Sánchez-García et al., 2022).

47
48 A considerable fleet of point-source methane-sensing satellites is now in orbit, including
49 purpose-built and repurposed instruments (Jacob et al., 2022). In the coming years, this number
50 will increase substantially (Jacob et al., 2022).

51
52 These satellites will have the greatest positive environmental impact if their results are widely
53 believed by a broad array of stakeholders across the world. Single-blind controlled methane
54 release testing, in which teams estimate methane emissions from one or more metered sources
55 without knowing the true rate, is an important and widely-used method of independently
56 determining the capabilities of a methane sensing system (Sherwin et al., 2021, 2023b; Bell et
57 al., 2020, 2022, 2023; Ravikumar et al., 2019; Rutherford et al., 2023).

58
59 In the first such single-blind release testing satellite systems, Sherwin et al. tested five satellites:
60 the commercial GHGSat-C and WorldView-3 systems and the government-supported PRISMA,
61 LandSat 8, and Sentinel-2 systems. In that study five teams analyzed data from different subsets
62 of these satellites (Sherwin et al., 2023b). This test demonstrated that, across the array of these
63 five satellites, this approach can be used to detect emissions ranging from 0.20 [95% confidence
64 interval = 0.19, 0.21] metric tons of methane per hour (henceforth t/h), for the most sensitive
65 systems, to 7.2 [6.8, 7.6] t/h. Relative quantification error was comparable to aircraft-based
66 methane sensing systems, although with significantly larger detection limits (Sherwin et al.,
67 2023b). Sample size was modest, however, with some satellites collecting only one
68 measurement, limiting generalizability of the results without additional data collection.

69
70 In addition, several methane-sensing satellites have launched since the previous test concluded in
71 2021, including the German EnMAP system and the 02 edition of the Chinese Gaofen 5
72 Advanced Hyperspectral Imager (GF5) and the 02E edition of the Ziyuan 1 Advanced
73 Hyperspectral Imager (ZY1) (EnMAP, 2023; Xinhua, 2022; Song et al., 2022). Although these
74 satellites were not primarily designed to sense methane, scientists have used similar systems to
75 detect substantial methane point sources from oil and natural gas infrastructure (Irakulis-Loitxate
76 et al., 2021).

77
78 This work conducts single-blind testing of nine distinct satellite systems, focusing on detection
79 and quantification performance for releases ranging from 0.03-1.6 t/h. In addition, we take steps
80 to evaluate the generalizability of our results through a highly transparent experimental design, in
81 which all teams submit full methane retrievals for the scene surrounding the release. This
82 approach provides insight into which identified methane emissions are unambiguously detected
83 and which might be difficult to distinguish from artifacts if the source location were not known.

84 **1 Materials and Methods**

85
86 We employed a fixed-location single-blind controlled methane release experimental design to
87 evaluate point-source methane sensing systems from October 10th to November 30th, 2022.
88 Participating teams were aware of the existence, timeframe, and precise location coordinates of
89 the test site. Teams were not informed during a given observation whether gas would be
90 released, nor of the size of released emissions. Teams were informed of an approximate upper
91 bound of 1.5 t/h. Teams were not given the precise configuration of Stanford equipment on the
92 ground, though large equipment may have been visible from space in some cases.

93
94 Metered controlled release volumes – including releases with zero volume – were retained by the
95 Stanford team and not given to teams until all estimates were submitted by all participants for all
96 stages of the test. Analysts estimated the presence and magnitude of methane emissions for each
97 overpass, with a reporting approach in compliance with the Advancing Development of
98 Emissions Detection (ADED) protocol for airplane and satellite systems (Zimmerle, 2022). More
99 information is provided in the Supporting Information (SI), Section S1.1.

100
101 We performed releases during overpasses of nine satellite constellations: the commercial
102 satellites GHGSat-C (GSC) of Canada and the US-based WorldView-3 (WV3), as well as

103 publicly-funded satellites, including the German Environmental Mapping and Analysis Program
104 (EnMAP), the Chinese Gaofen 5 (GF5), Ziyuan 1 (ZY1), and Huanjing 2 (HJ2), the Italian
105 PRecursores IperSpettrale della Missione Applicativa (PRISMA), the US LandSat (LS) 8 and 9,
106 and the pan-European Sentinel-2 (ESA, 2022a, b, c; Jervis et al., 2021; OHBI, 2022; EnMAP,
107 2023; Liu et al., 2019; USGS, 2022; Song et al., 2022; Zhong et al., 2021). With the exception of
108 the GHGSat-C constellation, none of these satellites was explicitly designed for methane
109 sensing, but their data have instead been applied to this end. Analysis teams first attempted to
110 estimate emissions volumes using available data from satellites and wind reanalysis products. In
111 some cases, multiple teams assessed the same observation from an instrument, providing an
112 opportunity to empirically assess variability due to source quantification algorithms, which
113 participating teams were not required to release. See the SI, Section S3 for the details each team
114 elected to share about their algorithms.

115
116 These satellites range from high-sensitivity/narrow swath to low-sensitivity/large swath, as
117 illustrated in Table 1. Revisit time is also anticorrelated with instrument sensitivity. The
118 Sentinel-2 and LandSat 8/9 systems have estimated detection limits of roughly 1-5 t/h (Gorroño
119 et al., 2023), but each satellite in these constellations covers the bulk of the world's landmass
120 every 10-16 days with a swath of 185-290 km (USGS, 2022; ESA, 2021a). GHGSat, EnMAP,
121 GF5, PRISMA, WorldView-3, and ZY1 are targeted "point-and-shoot" systems, with higher
122 resolution but narrower swaths of 12-60 km (ESA, 2022a, b; Jervis et al., 2021; OHBI, 2022;
123 EnMAP, 2023; Liu et al., 2019; Song et al., 2022). Existing publicly available information does
124 not specify whether HJ2 is targeted or has global coverage, but its swath of 800 km suggests it is
125 capable of global coverage (Zhong et al., 2021). Pixel size also varies widely across satellites,
126 with most tested satellites ranging from 20-30 m square pixels, while HJ2 has 6 km square pixels
127 and WorldView-3 has highly sensitive 3.7 m square pixels. Spectral resolution varies as well
128 across the tested satellites, from 0.3 nm for GHGSat-C and 200 nm for Sentinel-2 and LandSat
129 8/9 (Jacob et al., 2022), discussed further in the SI, Section S2. See the SI, Section S2 for
130 additional discussion of the capabilities of each satellite system.

1
2
3
4
5
6
7

Table 1. Key characteristics of each participating satellite constellation, from lowest to highest swath width, which is roughly proportional to an instrument’s minimum methane detection limit. Global coverage refers to a configuration that passively covers most of Earth’s surface over some number of orbits, while targeted coverage refers to a “point-and-shoot” instrument that must be pointed to a particular location. Nadir pixel size is presented here. Constellation size includes only active satellites. Accessing data from the GF5, ZY1, and HJ2 satellites requires permission from the Chinese government. Adapted with permission from (Sherwin et al., 2023b).

Satellite	Coverage	Constellation size	Swath [km]	Pixel size [m]	~Revisit time (per satellite)	Data availability	Source
GHGSat-C	Targeted	8 [§]	12	25x25	14 days	Commercial	(ESA, 2022a; Jervis et al., 2021)
WorldView-3	Targeted	1	13.1	3.7x3.7	1 day [‡]	Commercial	(ESA, 2022b)
PRISMA	Targeted	1	30	30x30	7 days	Public	(OHBI, 2022; ESA, 2012)
EnMAP	Targeted	1	30	30x30	4 days [†]	Public	(EnMAP, 2023)
Gaofen 5 (GF5)	Targeted	1	60	30x30	5-8 days [*]	Government	(Liu et al., 2019; Zhang et al., 2022; Luo et al., 2023)
Ziyuan 1 (ZY1)	Targeted	1	60	30x30	1-3 days [*]	Government	(Song et al., 2022)
Landsat 8/9	Global	2	185	30x30	16 days	Public	(USGS, 2022)
Sentinel-2	Global	2	290	20x20	10 days	Public	(ESA, 2021a)
Huanjing 2 (HJ2)	Unknown	2	800	6x6 km	≤4 days [*]	Government	(Zhong et al., 2021)

8 [§]Three of these GHGSat C satellites were launched after the conclusion of testing.
9 [‡]WorldView-3 requires a 4.5-day repetition cycle for best resolution within 20° off nadir.
10 [†]EnMAP requires a 27-day repetition cycle for best resolution within 30° off (Jacob et al., 2022).
11 ^{*}Revisit times for GF5, ZY1, and HJ2 are inferred, at least in part, from overpass schedules submitted by NJU.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

This test does not include the TROPOMI system on the Sentinel-5P satellite, which has a detection limit far above the maximum of the release apparatus used in this study (ESA, 2021b). We inquired about tasking the Earth Surface Mineral Dust Source Investigation (EMIT) satellite, launched by the US National Aeronautics and Space Administration (NASA) in July 2022 (Wang and Lee, 2022), but the system was not available to participate in this test.

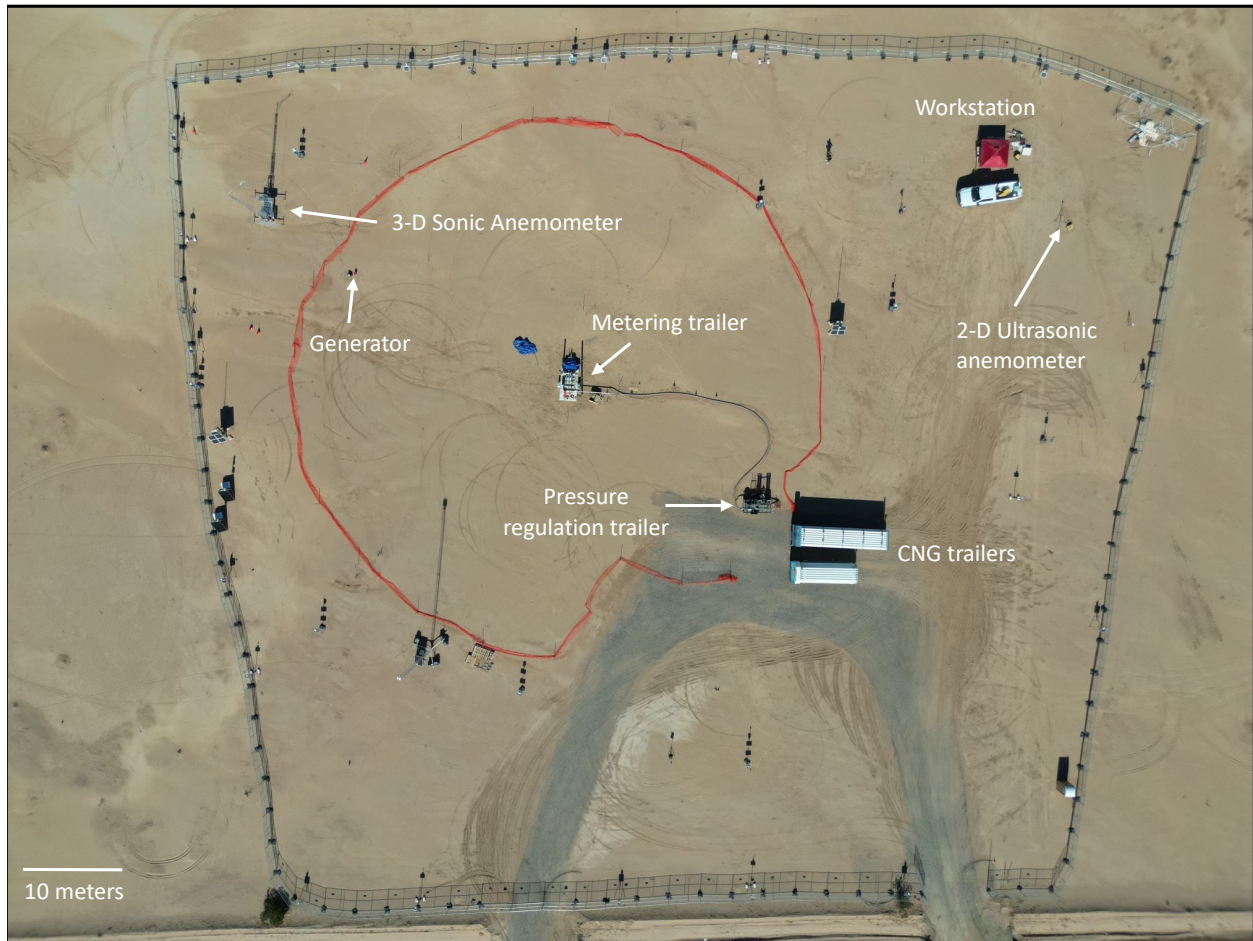
Participating analysis teams include private companies GHGSat (GHGSat, 2022), Kayrros (Kayrros, 2022), Maxar (Scott, 2022), and Orbio Earth (Orbio, 2023), as well as the Land and Atmosphere Remote Sensing (LARS) group of university researchers from Universitat Politècnica de València [Luís Guanter, Javier Roger Juan, and Javier Gorroño Viñegla (Irakulis-Loitxate et al., 2022a, b, 2021; Guanter et al., 2021)] and Nanjing University [Fei Li, Huilin Chen, and Yongguang Zhang (Jia et al., 2022)]. Each analysis team had the opportunity to submit estimates for all satellites tested, with the exception of the GHGSat-C satellites, to which GHGSat had sole access. See the SI, Section S3 for a description of each team and its members, as well as a list of invited teams that declined to participate.

19 **1.1 Materials**

20 For the full test period, our experimental equipment was located near Casa Grande, Arizona,
21 south of Phoenix, Arizona in the United States, with the release stacks located at [32.8218205°, -
22 111.7857730°].

23
24 The methane source was two trailers of compressed natural gas, shown in Figure 1, which passed
25 through a pressure regulation and reheating apparatus. The gas was then transmitted to the
26 metering and release trailer via a 7.62 cm (3 in) shipping hose at an exit pressure of roughly 150-
27 200 psig (1.03-1.37 Mpa), passing through one of three possible Coriolis meters before release
28 through one of two stacks, at a release height of either 7.3 or 3.0 m above ground level (El
29 Abbadi et al., 2023), shown in the SI, Figure 1. This testing setup approximately mimics an unlit
30 flare or tank vent on an oil and gas production site or other facility.

31
32
33



34
35 Figure 1. Aerial photograph of the site. Note that the workstation is ~60 m from the release apparatus and ~50m
36 from the compressed natural gas (CNG) trailers. Reproduced with permission from (El Abbadi et al., 2023).

37 This experiment was designed to provide near-optimal conditions for methane-sensing satellites.
38 In addition to the desert background, the site contained only equipment necessary to conduct
39 controlled methane releases and test a suite of methane sensing technologies. The result is a

40 significantly less complex scene than many oil and gas facilities, which will often contain
41 multiple pieces of infrastructure such as wellheads, tanks, flares, and separators at production
42 sites, and entire buildings with sophisticated machinery and piping at compressor stations and
43 gas processing plants. More complex scenery can make methane remote sensing more
44 challenging. Future work with scenes that more closely mimic industrial sites will help determine
45 the associated differences in technology efficacy, if any.

46
47 Achievable release rates for the three Coriolis meters, installed in pipes of different diameter,
48 were 2 – 30 kilograms per hour (kg/h), 30 – 300 kg/h, and 300 – 2,000 kg/h for natural gas. See
49 (El Abbadi et al., 2023) for further detail.

51 **1.2 Safety**

52 All natural gas equipment fabrication, operation, and transportation was conducted by personnel
53 affiliated with Rawhide Leasing, a gas services contractor. Stanford personnel contributed to
54 assembly of some equipment, but did not operate natural gas release equipment or pass within
55 our 100-foot (30.5 m) safety perimeter fence during active releases. The research workstation,
56 from which Stanford researchers coordinated data collection and related field operations, was
57 ~60 m away from any equipment through which natural gas flowed.

58
59 In addition, Stanford researchers periodically monitored plume dissipation in real time via a
60 FLIR GasFinder 320 infrared camera and continuously paid attention to olfactory signals from
61 the gas, which was odorized. The infrared camera showed clearly that the plume dissipated well
62 before reaching any on-site personnel. Equipment design contributed to this intrinsic safety,
63 because the emission source was elevated off the ground and gas often exited at a high vertical
64 velocity, particularly at larger release volumes, accelerating natural methane lofting. When
65 Stanford researchers detected gas smell during testing, they diligently checked infrared footage
66 of the plume and/or ambient wind conditions to ensure safety of all personnel onsite.

68 **1.3 Data logging**

69 Stanford researchers collected data logs directly from the Coriolis gas flow meters, accounting
70 for modest timestamp offsets as described in (El Abbadi et al., 2023).

72 **1.4 Data collection procedures**

73 All satellite-coincident releases began at least 15 minutes before the scheduled satellite overpass
74 time, provided by participating teams.

75
76 Stanford personnel set all release levels remotely, using WiFi-enabled control software deployed
77 on a laptop computer. For releases conducted on or before October 20th, Stanford personnel set a
78 desired flow rate, with an automated control system adjusting valves in real-time to target that
79 rate. After it became clear that this approach resulted in unnecessary flow rate variability,
80 releases from October 21st on were conducted by setting the relevant valve to a desired level of
81 openness, improving flow stability while slightly reducing the system's ability to target a specific
82 release rate, although this system still represents a major improvement over the manual approach
83 employed in (Sherwin et al., 2023b). Flow can fluctuate during the releases due to shifts in
84 pressure, temperature, and simple turbulent flow through the system. All performed releases

85 except four had flow variability with a 5-minute 95% confidence interval within $\pm 10\%$ of mean
86 flow. On November 15th, a GF5 satellite acquisition was rescheduled without notice to the
87 Stanford team for a time that happened to be one minute after conclusion of a different satellite
88 release, resulting in flow variability within $\pm 20\%$ of the 5-minute mean. Three additional
89 releases exceeded a 5-minute flow variability 95% confidence interval of $\pm 10\%$: the October 11th
90 GHGSat-C overpass (in which the instrument was not tasked), the October 17th WorldView-3
91 release of 0.042 [0.034, 0.050] t/h, and the November 30th PRISMA release of 0.98 [0.87, 1.08]
92 t/h.

93
94 Interference from other sources was examined and found to be minimal. Over the course of the
95 experiment, we tested the Carbon Mapper, GHGSat AV, Kairos Aerospace, MethaneAIR, and
96 Scientific Aviation aerial methane sensing systems (El Abbadi et al., 2023), all of which are
97 more sensitive than any of the satellites tested. These aircraft, which also surveyed the nearby
98 area during the process of data collection, found no detectable methane sources outside our test
99 site. This strongly suggests that our test was free of interference from significant confounding
100 methane sources. The only evidence of modest possible landfill interference comes Scientific
101 Aviation, whose highly sensitive in situ measurement technology found modest and diffuse
102 methane concentration enhancements over a nearby landfill, potentially impacting only one of
103 the three days of testing, and only one of the seven measurements conducted on that day (El
104 Abbadi et al., 2023).

106 **1.5 Flow rate uncertainty**

107 Sources of uncertainty in measured methane flow rates include variability in actual natural gas
108 flow rates (represented as the standard deviation of metered natural gas flow over a 5-minute
109 period), rated meter uncertainty, and uncertainty in gas composition, which can vary even for a
110 consistent supplier. We used highly precise Coriolis meters, which have manufacturer rated
111 uncertainty of 0.25% of the flow rates used in this study (El Abbadi et al., 2023). Natural gas
112 composition for the gas used in these releases, derived from measurement stations on the
113 transmission pipeline that supplied the gas used in this test, ranged between 93.6% [93.3%,
114 93.9%] and 95.4% [94.7%, 96.1%] methane, described further in the SI, Section S1.2 and in
115 reference (El Abbadi et al., 2023). We propagate these sources of error into our metered values
116 using code listed in data and code availability statement. See (El Abbadi et al., 2023) for further
117 discussion of sources of metering uncertainty and our method of determining flow rate
118 uncertainty, as well as detailed gas composition data.

119
120 Following (Sherwin et al., 2023b), we use a 5-minute averaging period used to compute flow
121 variability. This is based on the fact that a plume traveling with a relatively slow average wind
122 speed of 2 m/s, the minimum observed 5-minute average wind speed for any valid satellite
123 measurement, would traverse 600 m within 5 minutes (300 seconds). By this distance, much of
124 the originally emitted methane has likely dissipated into background concentrations, with the
125 bulk of the methane enhancement detected by a satellite remaining closer to the release point.

127 **1.6 Experimental design**

128 This single-blind field trial employed a two-stage experimental design, modeled on (Sherwin et
129 al., 2023b). This approach aims to disentangle the effect of wind speed uncertainty from other
130 sources of methane quantification uncertainty, e.g. due to algorithmic differences.

131
132 Stanford personnel released metered quantities of methane from the test site via procedures
133 described above and in reference (El Abbadi et al., 2023). The Stanford ground team and
134 contract personnel operating equipment communicated no information to participating teams
135 regarding metered flow rates or metered wind speed or direction. Participating teams were aware
136 of the precise location coordinates of the test, but were not informed of the precise configuration
137 of ground-based equipment within the test site. Teams were given a rough range of possible
138 overall flow rates, from below 0.01 t/h to roughly 1.5 t/h. To facilitate efficient tasking of
139 government satellites, LARS and NJU were informed in advance that weekend releases in
140 November would be cancelled and all such dates were excluded from single-blind analysis for
141 those teams. In addition, participating teams were not informed of the details of the equipment or
142 its configuration, or the diameter of the pipes and hoses involved, although teams were informed
143 that the test would use compressed natural gas as the methane source.

144
145 After each team submitted final stage 1 estimates based on the above information, we proceeded
146 to stage 2 estimates. In stage 2, Stanford provided 10 m wind speed and direction data from our
147 on-site ultrasonic anemometer (shown in Figure 1) at one-second resolution and teams were
148 allowed to re-estimate emissions based on measured ground wind conditions rather than re-
149 analysis products as in stage 1. All teams submitted stage 1 and stage 2 estimates, with the
150 respective timelines described in the SI, Section S2.10. Note that turnaround time for results in
151 this study may not be representative of commercial or field performance.

152

153

154 **2 All tested satellites detected methane**

155

156 For the eight satellites given nonzero methane emissions, at least one analysis team correctly
157 detected methane. The single HJ2 measurement, using the HJ2B satellite, was rescheduled
158 without notice to a time in which Stanford was not releasing methane.

159

160 In total, the nine tested satellites conducted 82 overpasses. Six analysis teams analyzed data from
161 between 1 – 8 satellites each, resulting in a total of 492 potential estimates. Stanford filtered
162 many of these estimates from analysis before teams submitted results, for various reasons (e.g.
163 due to release system malfunction or prior notice to teams tasking government satellites that
164 there would be no weekend releases in November). In addition, most teams opted to submit
165 estimates for only a subset of all available satellites. See the SI, Section S1.3 for further
166 discussion of data exclusion criteria.

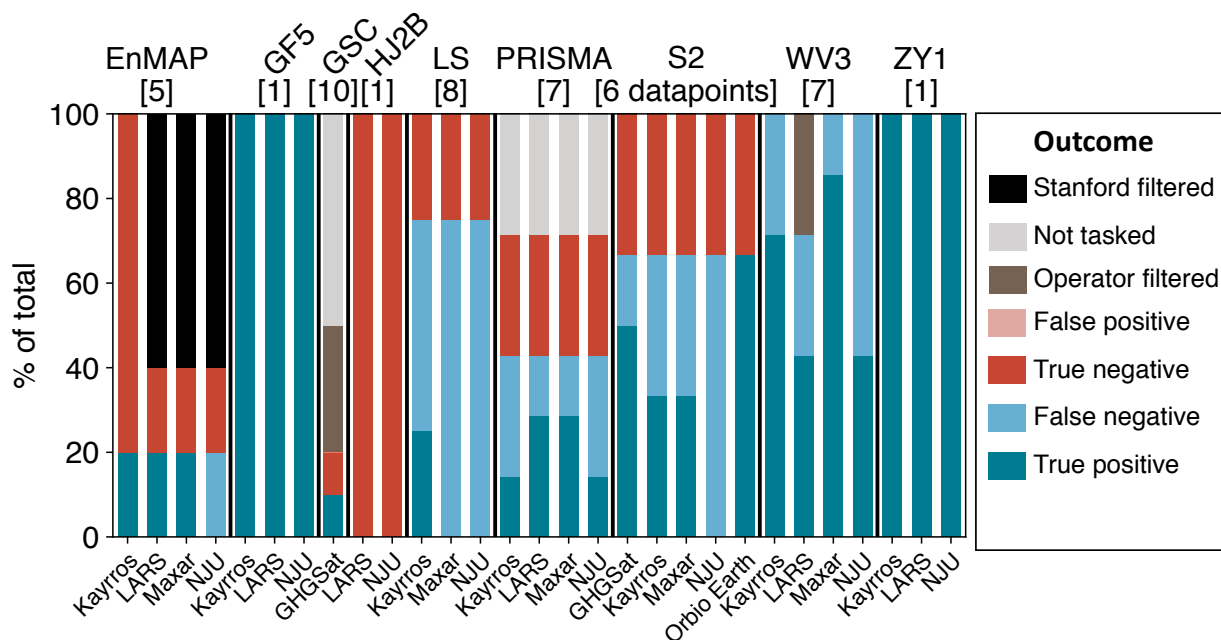
167

168 Of the 139 estimates not filtered by Stanford, in five instances (3.6% of the total), teams filtered
169 estimates using internal quality control criteria related to cloud cover, image clipping, or other
170 factors that could compromise the ability to produce a valid methane estimate. GHGSat filtered
171 three retrievals from the GHGSat-C satellite due to clouds (see Figure 8 and the SI, Section S4
172 for sky images and further discussion of clouds). LARS filtered two WorldView-3 retrievals due
173 to cloud cover (November 22th) and inconsistent wind, and possible effects of human-made
174 surface features (October 10th). As a result, a total of 134 estimates included valid methane
175 detection estimates.

176

177 Of these 134 estimates, 80 (58%) were identified as either a true positive or true negative,
178 correctly determining the presence or absence of methane, as shown in Figure 2. True positives
179 represent 46 (34%) of total estimates with valid detection estimates, with 34 (25%) true
180 negatives. Note that for Sentinel-2, we consider non-detection of an 0.005 t/hr release on
181 November 28th to be a true negative, as this value is more than two orders of magnitude below
182 existing estimates of the detection threshold of this system (Gorroño et al., 2023; Sherwin et al.,
183 2023b).

184



186 Figure 2. Detection performance by satellite and team. The total number of measurements per satellite is listed in
 187 brackets, excluding measurements filtered by Stanford across all teams. All teams analyzing data from the three
 188 Chinese satellites, Gaofen 5 Advanced Hyperspectral Imager (GF5), Ziyuan 1 (ZY1), and Huanjing 2B (HJ2B) all
 189 correctly classified all emissions. Detection performance varied substantially across the Sentinel-2 (S2) and LandSat
 190 8/9 (LS) wide-area satellites. On several days, anticipated measurements from PRISMA and GHGSat-C (GSC) were
 191 not collected because the satellite was not tasked. In others, e.g. two WorldView-3 retrievals from LARS, no
 192 retrieval was conducted due to concerns over image clipping or excessive cloud cover. No teams submitted false
 193 positives, in which they reported the presence of methane when none was released.
 194

195 Of the 41 false negatives (30%), most (25) are concentrated in the lower-sensitivity Sentinel-2
 196 and LandSat 8/9 systems. There is substantial variability in false negative rates across teams. For
 197 example, Orbio Earth correctly classifying all valid Sentinel-2 releases. GHGSat missed only one
 198 Sentinel-2 release, and NJU detected none. This highlights that analysis of identical spectral data
 199 can produce very different results. As in (Sherwin et al., 2023b), there were no false positives,
 200 defined as incorrect reports of the presence of methane.
 201

202 In several cases, a satellite was not tasked during an overpass for which the Stanford team
 203 conducted a release, either due to technical issues, scheduling issues, or miscommunications
 204 between the Stanford team and the operator. This occurred for five GHGSat overpasses and two
 205 PRISMA overpasses, resulting in a total of 13 Not Tasked estimates from participating teams for
 206 these two satellites, 9% of all estimates not filtered by Stanford.
 207

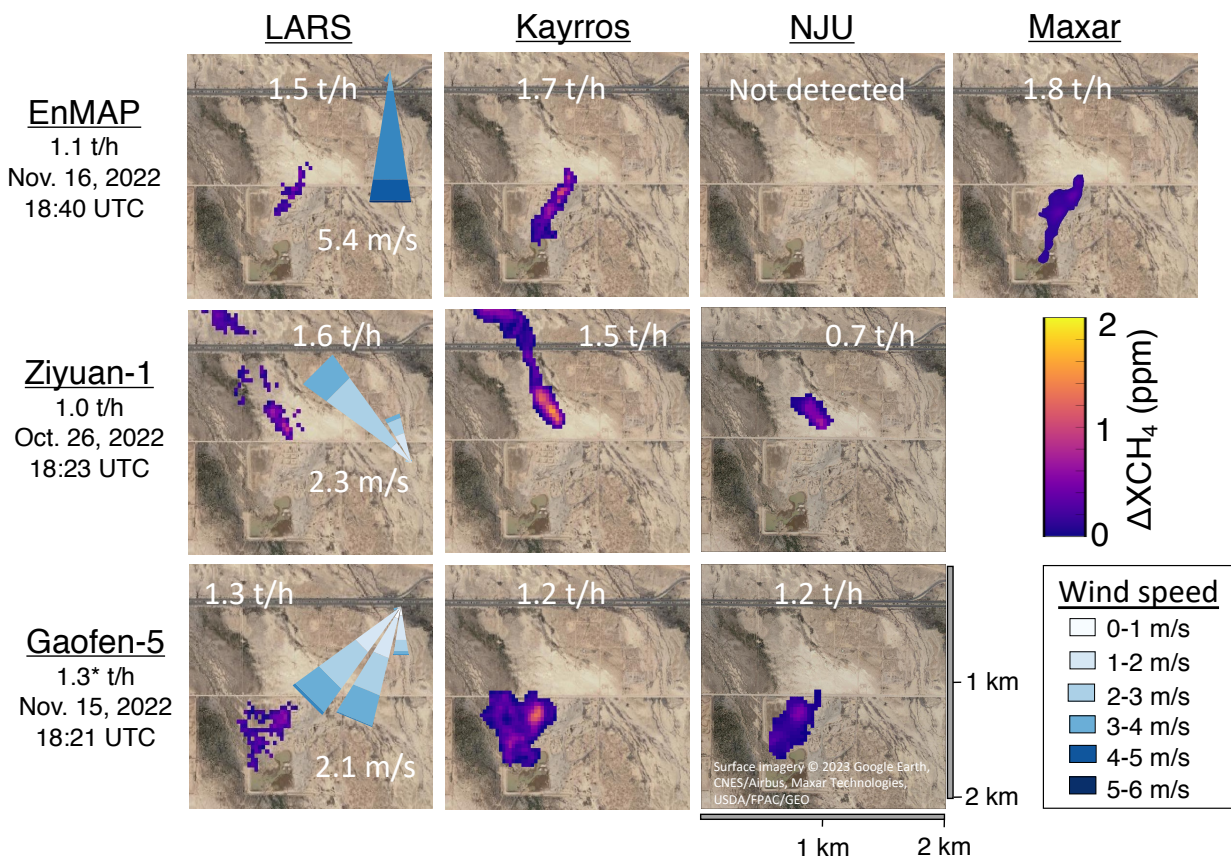
208 2.1 First-time single-blind detections from three satellites

209 This work includes the first-ever single-blind test of the Chinese Ziyuan 1 (ZY1), Gaofen 5
 210 (GF5), and Huanjing 2B satellites (HJ2B), as well as the European EnMAP satellite. Previous
 211 studies have used a subset of these satellites to detect and quantify point-source emissions with
 212 estimated magnitudes as small as 0.5 t/h, but have not performed ground-truth testing (Irakulis-
 213 Loitxate et al., 2021). Roger et al. compare EnMAP retrievals with the single-blind-validated

214 PRISMA satellite as a benchmark, finding promising results, especially for offshore emissions of
 215 1 t/h or more (Roger et al., 2023).

216
 217 Figure 3 shows masked methane plume images from ZY1, GF5, and EnMAP, over a standard
 218 optical satellite image background, for emissions of roughly 1 t/h. Masking refers to the process
 219 of spatially differentiating a methane emission from background noise. The HJ2B acquisition
 220 was rescheduled without prior notice to the Stanford team to a time at which no release took
 221 place, which all teams analyzing HJ2B data correctly identified as a non-emission. We present
 222 images from all teams analyzing satellite data from these measurements, including LARS,
 223 Kayrros, NJU, and Maxar. See the SI, Section S4 for masked and unmasked plume images for all
 224 satellites and teams.

225
 226



227
 228 Figure 3. Visualization of detected emissions for the newly-tested European and Chinese satellites, using the release
 229 closest to 1 t/h in all cases. The true measured emission rate, as well as the timestamp are shown below the satellite
 230 name. Mean estimated volume from each team/satellite pair, as well as a 5-minute wind rose of measured 10-meter
 231 wind speed and the direction toward which the wind was blowing, are superimposed on the corresponding picture.
 232 The wind rose represents a histogram of one-second wind measurements in each direction, broken down by wind
 233 speed. Where an emission was not detected, we show the full unmasked retrieval field. Cloud-free surface imagery
 234 © 2023 Google Earth, CNES/Airbus, Maxar Technologies, USDA/FPAC/GEO. *The Gaofen 5 measurement was
 235 rescheduled without notice to a time that happened to be one minute after releases had concluded for a different
 236 satellite, resulting in artificially high variability in the metered ground-truth flow rate.

237 Note that, as was observed in (Sherwin et al., 2023b), teams analyzing precisely the same
 238 spectral data can produce methane plume masks with very different shapes. Each row represents

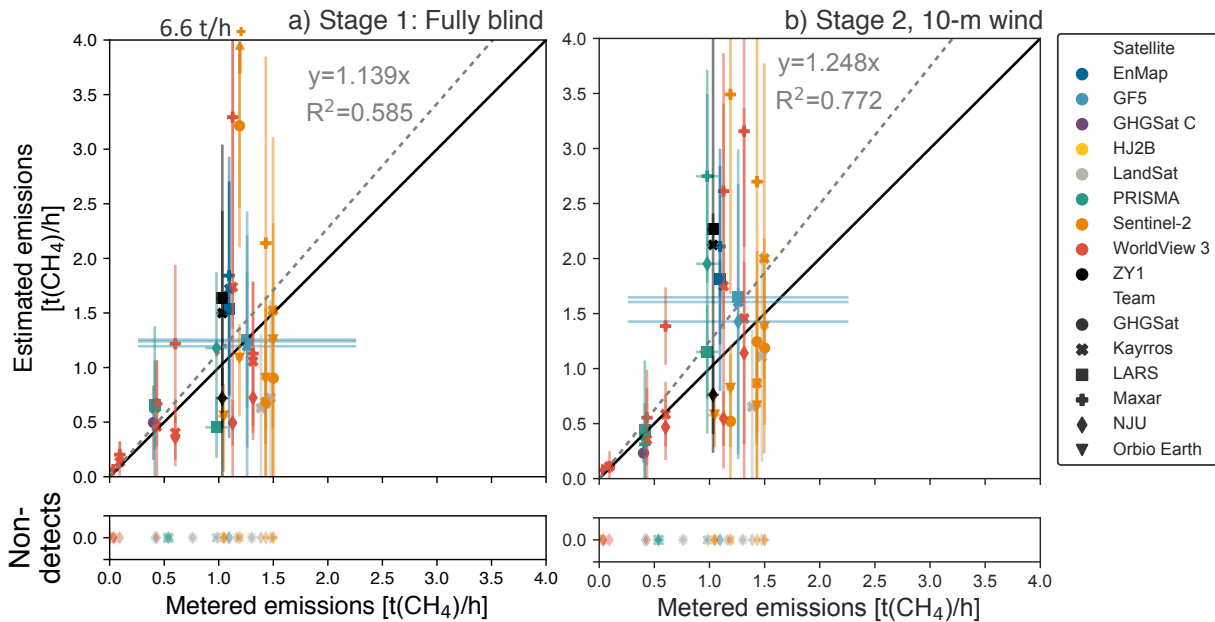
239 a distinct satellite, while each column shows estimates from a distinct team. For example, the
240 first row shows estimates for the November 16th EnMAP satellite measurement, for which four
241 teams submitted estimates. Three of the four teams detected the emission. LARS, Kayrros, and
242 Maxar all show masked plumes traveling in roughly the same direction, but the Kayrros and
243 Maxar plumes are fairly contiguous, while the LARS plume is smaller and contains disjunct or
244 tenuously-connected clusters of estimated methane enhancements. Overall, masks from LARS
245 are more conservative and less spatially contiguous than other teams. However, quantification
246 estimates from LARS, Kayrros, and Maxar all have overlapping quantification intervals,
247 demonstrating that the results are not statistically distinguishable across these three teams (NJU
248 did not detect this EnMAP emission). Even with cases with large mean differences, e.g. October
249 26th estimates for ZY1, which range from 1.6 [1.2, 2.0] t/h for LARS to 0.7 [0.6, 0.9] t/h for
250 NJU, the 95% confidence intervals overlap. These findings suggest that many factors influence
251 quantification performance, even when working with identical spectral data, but large
252 uncertainties make disentangling these differences a challenge. Further analysis of these
253 algorithmic differences is beyond the scope of this work, as teams were not asked to provide
254 algorithmic details, which are often proprietary. Further experimentation may enable analysis of
255 general trends in advantages of one algorithm over another, but the order-100 number of
256 datapoints here is insufficient to make such judgements.

257
258 Wind can vary substantially in speed and direction even on five-minute timescales relevant to
259 methane quantification, as shown in wind roses inset in the left-most panel for each satellite in
260 Figure 3. This variability clearly influences plume formation, with emissions with steadier wind
261 directions and higher speed, such as the EnMAP and ZY1 measurements shown here (5.4 [3.7,
262 7.2] m/s and 2.3 [1.0, 3.7] m/s average wind speed, with a wind direction circular standard
263 deviation of 16° and 11°, respectively), resulting in narrower plumes. The highlighted GF5
264 measurement has slower and more variable winds and a wider plume in all three retrievals (2.1
265 [0.3, 4.0] m/s, with a wind direction circular standard deviation of 18°).

267 **2.2 Reliable overall quantification performance**

268
269 Releases in this study covered a wide range of emission rates, as low as 0.0332 [0.0328, 0.0336]
270 t/h, analogous to a medium-sized liquids unloading event at an oil and gas production site (Bell
271 et al., 2017), and as high as 1.48 [1.43, 1.52] t/h, analogous to a medium-sized unlit flare
272 (Cusworth et al., 2021). For all detected emissions, mean estimates for all satellite-team
273 combinations are between -56% and 456% of the metered value (Figure 4; see also SI, Section
274 S5), with 55% of nonzero estimates falling within $\pm 50\%$ of the metered value. Excluding
275 estimates from Maxar, which discovered after submitting results that its estimates were likely a
276 factor of 2.3 too high due to a misinterpretation of a deprecated spectral absorption library, this
277 fraction rises to 63% (Hayden and Christy, 2023). However, the best-fit line across all satellite
278 measurements, any one of which may have substantial quantification error, is largely unbiased,
279 with a slope close to the ideal value of 1 (which would denote perfect agreement on average).

280
281



283
 284 Figure 4. Methane quantification performance by satellite and team. Metered emissions compared with single-blind
 285 estimates for each overpass with successfully reported data, with 95% X and Y confidence intervals. a) Fully blind
 286 stage 1 results using modeled wind speed estimates. Note one Sentinel-2 estimate exceeds the y-axis limit at 6.6
 287 t(CH₄)/h. b) Stage 2 results using on-site 10 m wind speed and direction measurements. LARS WorldView-3
 288 quantification estimates are excluded from the main analysis, as stage 1 estimates were submitted after wind data
 289 had been unblinded to a member of the LARS team not involved in analyzing WorldView-3 data, while
 290 corresponding stage 2 estimates were submitted after release volumes were unblinded. Note that Maxar submitted
 291 PRISMA estimates for stage 2 only. The grey dashed lines represent an ordinary least squares fit with the intercept
 292 fixed at zero, with slope and uncentered R² displayed. Maxar has since determined that its estimates were likely
 293 artificially high, potentially introducing upward bias into aggregate statistics (Hayden and Christy, 2023). See the SI,
 294 Section S4.2 for a version of this plot excluding Maxar, which shows overall improvement in both slope and R². The
 295 black solid lines denote exact 1:1 agreement. See the SI, Section S4 for satellite- and team-specific results.

296 In percent quantification error terms, this overall performance approaches that of the satellites
 297 and teams tested in Sherwin et al. 2023, in which 75% of estimates fell within $\pm 50\%$ of the
 298 metered value, demonstrating a relative error profile similar to that observed in aircraft-based
 299 methane remote sensing technologies (albeit with minimum detection limits one to three orders
 300 of magnitude larger) (Sherwin et al., 2023b; El Abbadi et al., 2023; Bell et al., 2022). Direct
 301 comparison with the results in Sherwin et al. 2023 is complicated by the fact that releases in this
 302 study focused on smaller emissions, with a maximum of roughly 1.5 t/h instead of 7.2 t/h.
 303 Aircraft-based methane remote sensing technologies tested in El Abbadi et al. tend to have
 304 modestly better quantification performance in percentage terms, with 68-80% of estimates from
 305 Carbon Mapper, GHGSat, Kairos Aerospace, and MethaneAIR falling within $\pm 50\%$ of the
 306 metered value (El Abbadi et al., 2023), a substantial improvement over prior tests of the same
 307 technologies (Sherwin et al., 2021; Rutherford et al., 2023). In each of these cases, best-fit lines
 308 have a slope that is similarly close to 1:1 agreement.

309
 310 See the SI, Section S4 for error summary statistics by satellite and team. Error bars in metered
 311 values along the x-axis are generally too small to be visible, with the notable exception of the

312 GF5 measurement, which was rescheduled without notice to a time that happened to be one
313 minute after releases had concluded for a different satellite.

314
315 In stage 2 of the test, teams produced updated results using measured 10 m wind data from an
316 on-site three-dimensional ultrasonic anemometer, though still blind to released volumes.
317 Applying an ordinary least squares linear fit to all quantified emissions, with the intercept set to
318 zero, we see a modest increase in slope, rising from 1.139 [0.832, 1.446] in stage 1 to 1.248
319 [1.037, 1.459] in stage 2 (Figure 4).

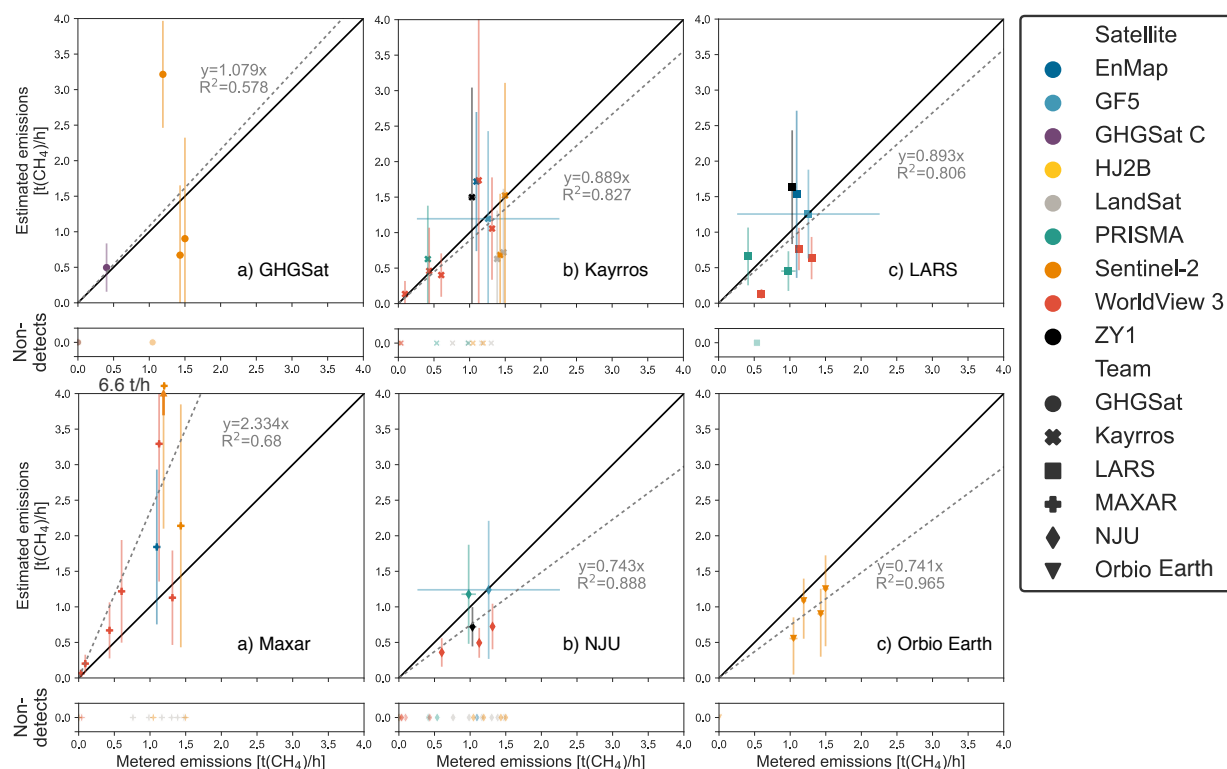
320
321 Interpretation of these results is complicated by the fact that the Maxar team discovered after
322 submitting blinded results that the spectral library underlying their estimates contained an error
323 that likely artificially inflated their estimates by a factor of 2.3, discussed in detail in a white
324 paper produced by Maxar personnel (Hayden and Christy, 2023). This is consistent with the
325 Maxar-specific parity chart in the SI, Section S4, alongside other satellite- and team-specific
326 results, which shows a regression best fit line of 2.334 [1.030, 3.638] and an uncentered R^2 of
327 0.96, indicating a close linear fit. Excluding Maxar results (as in the SI, Section S4.2), the Stage
328 1 slope for all remaining teams falls to 0.897 [0.716, 1.078], with a Stage 2 slope of 1.010
329 [0.841, 1.180], almost perfect average agreement with metered values. These slopes are 21% and
330 19% below the respective estimates in which Maxar values were included.

331
332 Note that LARS WorldView-3 quantification estimates are excluded from the main analysis, as
333 stage 1 estimates were submitted after wind data had been unblinded to a member of the LARS
334 team not involved in analyzing WorldView-3 data, while corresponding stage 2 estimates were
335 submitted after release volumes were unblinded. Although the Stanford team believes all LARS
336 quantification estimates for WorldView-3 were submitted without leveraging unblinded data, we
337 must exclude them from the main analysis. This does not affect the integrity of detection
338 estimates, as only wind measurements were unblinded when these were first submitted. See the
339 SI, Section S4 for LARS WorldView-3 quantification results.

340
341 After incorporating on-site wind measurements, the uncentered R^2 increases from 0.585 to 0.772,
342 a substantial improvement in goodness-of-fit. Excluding Maxar results, these numbers rise to
343 0.768 and 0.826, respectively. The linear fit presented here treats all estimated emission rates
344 from all team as independent datapoints. Note that uncentered R^2 values from such a linear fit,
345 with a zero intercept, have a different interpretation than R^2 values from nonzero-intercept
346 regressions and should not be compared directly. See (Sherwin et al., 2023b), SI Section S5 for
347 further explanation of the reasons for an ordinary least squares fit with the intercept fixed to zero.
348 This improved average linear fit with in situ wind does not necessarily translate to lower error for
349 each individual satellite, as shown in the SI, Section S4, alongside additional regression results.

350
351 Confidence intervals submitted by teams appear to be modestly overconfident. For Stage 1
352 estimates, the metered value is within the provided 95% confidence interval only 70% of the
353 time, somewhat below the expected value of 95% for perfectly-calibrated 95% confidence
354 intervals. For Stage 2, this fraction falls to 52%, although mean error improves. Note that these
355 values combine results from multiple satellites and teams, and thus represent an overall sense of
356 the performance of satellite-based methane sensing systems as a technology class. Additional
357 data collection is needed to characterize the performance of each individual satellite in detail.

358
359 Figure 5 shows Stage 1 fully blinded results, the same underlying data as in Figure 4, for each
360 individual team. Team-specific parity lines tend to fall near the ideal 1:1 level, with Orbio Earth
361 and NJU exhibiting modest low bias parity slopes of 0.74. Note that Maxar's parity slope of 2.3
362 matches almost exactly with the factor of 2.3 they believe was error introduced into their system
363 through misinterpretation of a deprecated spectral library (Hayden and Christy, 2023). The bulk
364 of false negatives were from the relatively low-resolution Sentinel-2 and LandSat 8/9 satellites.
365 However, Orbio Earth successfully detected all Sentinel-2 releases, except a release below 0.010
366 t/h (testing another technology), far below all estimates of the Sentinel-2 detection limit
367 (Gorroño et al., 2023; Sherwin et al., 2023b). These results highlight algorithmic variation across
368 teams analyzing the same spectral data.
369
370



371
 372 Figure 5. Parity charts by team, for fully blinded Stage 1 estimates only. Metered emissions compared with single-
 373 blind estimates for each overpass with successfully reported data, with 95% X and Y confidence intervals. Note one
 374 Maxar Sentinel-2 estimate exceeds the y-axis limit at 6.6 t(CH₄)/h. LARS stage 1 WorldView-3 quantification
 375 estimates are excluded from the main analysis, as they were submitted after wind data had been unblinded to a
 376 member of the LARS team not involved in analyzing WorldView-3 data. The grey dashed lines represent an
 377 ordinary least squares fit with the intercept fixed at zero, with slope and uncentered R² displayed. Maxar has since
 378 determined that its estimates were likely artificially high, potentially introducing upward bias into aggregate
 379 statistics (Hayden and Christy, 2023). The black solid lines denote exact 1:1 agreement. See the SI, Section S4 for
 380 Stage 1 and Stage 2 satellite- and team-specific results.

381

382 2.3 Qualitatively assessing detection performance in the field

383

384 The smallest emission detected by each team gives a rough upper bound on the lower detection
 385 capabilities of each instrument, at least in a desert environment with a known release location.
 386 We compare these smallest detected emissions with previous estimates of lower detection
 387 capabilities of each satellite. The smallest emission detected was 0.0332 [0.0328, 0.0336] t/h,
 388 identified by Maxar using WorldView-3, shown in Figure 6. Kayrros also detected an emission
 389 below 0.1 t/h using WorldView-3. This is consistent with previous estimates of lower detection
 390 capabilities, with Sánchez-García et al. detecting an emission estimated at ~0.040 t/hr in
 391 Turkmenistan using WorldView-3 (Sánchez-García et al., 2022).

392

393 Orbio Earth, Maxar, and GHGSat all detected a 1.19 [1.15, 1.23] t/h emission using Sentinel-2,
 394 with errors ranging from -8% to +170%. Orbio Earth detected a 1.05 [0.99, 1.10] t/h emission to
 395 within ±47%. These emissions are 15-25% below the smallest emission detected using Sentinel-
 396 2 in any previous satellite controlled methane release test, and consistent with simulation-based

397 estimates (Sherwin et al., 2023b; Gorroño et al., 2023). The story is similar for LandSat 8/9, with
398 the smallest detected emission at 1.39 [1.34, 1.43] t/h. This is also slightly below estimated lower
399 detection capabilities in the literature (Jacob et al., 2022).

400
401 The smallest emission detected via PRISMA was 0.414 [0.410, 0.417] t/h smaller than the 0.5-
402 2.0 t/h estimated by Guanter et al. as PRISMA's lower detection threshold (Guanter et al., 2021).
403 The smallest detected emissions for the remaining satellites are 1.10 [1.06, 1.13] t/h for EnMAP,
404 1.26 [0.26, 2.26] t/h for GF5, and 1.03 [0.98, 1.09] t/h for ZY1. However, given that the
405 technical characteristics of these three satellites are similar to PRISMA, they can likely be used
406 to detect emissions below 1 t/h, at least under favorable environmental conditions (Jacob et al.,
407 2022; Roger et al., 2023).

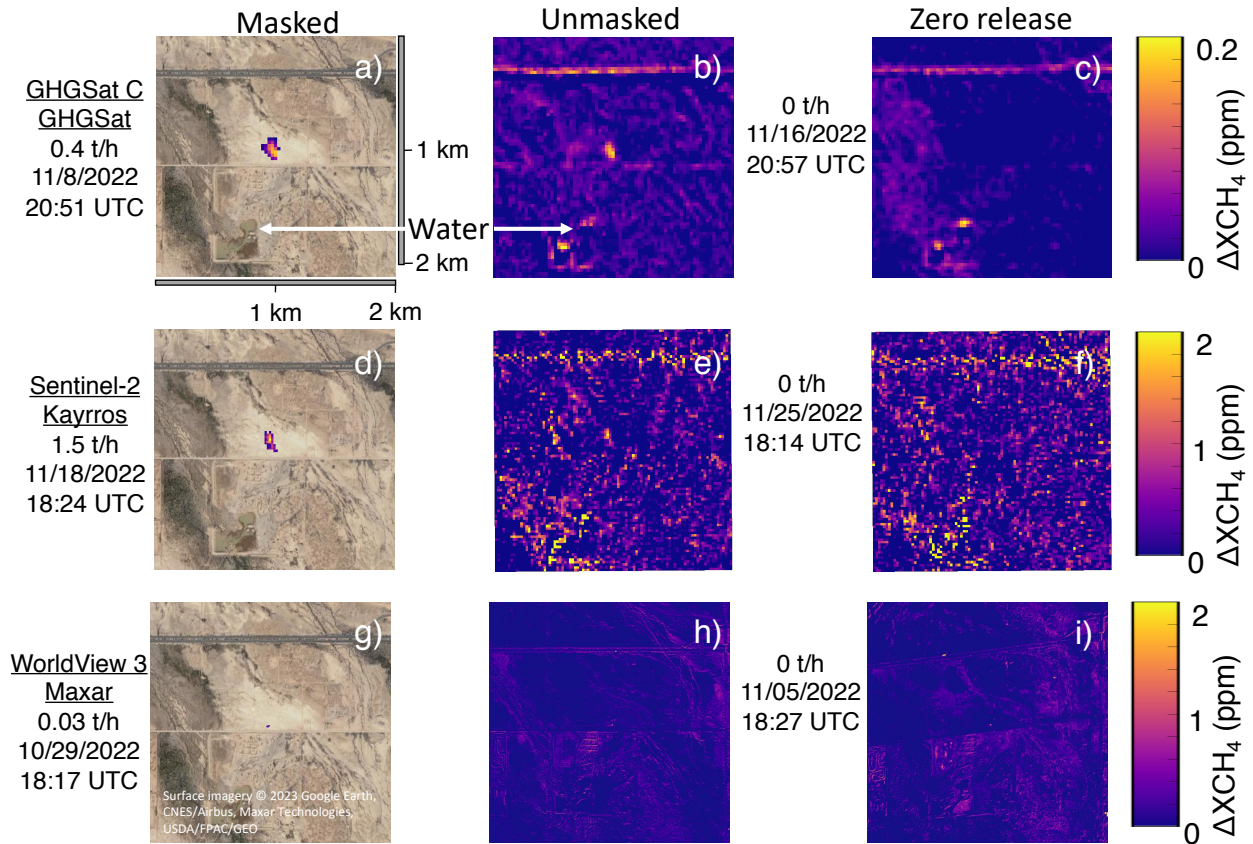
408
409 GHGSat correctly detected and quantified the only nonzero release for which GHGSat-C
410 collected data and passed quality control, which was 0.401 [0.399, 0.404] t/h, roughly double the
411 smallest release GHGSat quantified using the same satellite system in (Sherwin et al., 2023b).
412 GHGSat's lower detection threshold is estimated at 0.1-0.2 t/h (Jacob et al., 2022). HJ2B was not
413 tasked during any active releases, meaning that future testing is needed to assess its detection
414 capabilities.

415
416 In practical applications for global remote sensing, teams have only limited information about
417 the location of possible sources and their likelihood of emitting at visible levels. As a result, it is
418 possible that the known-location experimental design applied here may have allowed teams to
419 artificially boost detection sensitivity to levels that would be difficult to achieve in general
420 practice.

421
422 To qualitatively assess this possibility, all teams were required to submit methane retrieval field
423 images for all submitted estimates, including both detections and non-detections. In all cases,
424 teams submitted full-scene retrieval fields in a 2x2 km box around the release location. For
425 detected emissions, teams also submitted masked plume images, overlaying the estimated
426 methane plume above an optical image of the background location. See the SI, Section S4 for all
427 such images.

428
429 We highlight selected images in Figure 6 to showcase issues related to spectral artifacts, e.g.
430 apparent methane enhancements due to water bodies, clouds, or roads, that we were not able to
431 quantitatively address in this study. The GHGSat images, shown at a contrast-enhancing
432 narrower color scale of 0-0.2 ppm instead of this study's standard 2 ppm, show that for the
433 November 8th retrieval of the 0.401 [0.399, 0.404] t/h release, there are pixel clusters with
434 enhancements of comparable magnitude outside of the release area. However, these
435 enhancements are concentrated along ground features such as a water body southwest of the site
436 and a highway north of the site, confirmed in Google Maps imagery and WorldView-3 optical
437 images in the SI, Section S4. As a result, automated or manual intercomparison of the spatial
438 overlap of apparent methane enhancements and ground features visible in optical imagery could
439 plausibly help differentiate between such signal artifacts and true emissions. In some cases, it
440 may be possible to use measurements in which there is no evidence of a methane emission, e.g.
441 the November 16th measurement (in which GHGSat correctly determined the absence of methane
442 in a single-blind manner), to gain additional information into ambiguous cases. Artifacts such as

443 the water feature may consistently appear across retrievals, which could suggest that they are not
 444 true methane enhancements. Furthermore, GHGSat flagged the water body in both retrievals as a
 445 potential artifact, indicating that it would likely have been possible to correctly identify only the
 446 true methane emission in the November 8th scene even without a reference image with no
 447 methane.
 448



449 Figure 6. Masked and unmasked retrievals for selected emissions. In each case, the unmasked retrieval in the middle
 450 column appears to contain artifacts of similar intensity and shape to the masked emission. However, the emission
 451 may be more distinguishable from artifacts after intercomparison with ground features revealed through optical
 452 imagery, e.g. the water body southwest of the release site, and intercomparison with a reference day with zero
 453 emissions, as in the right column. Note that the GHGSat retrievals use a higher-contrast scale of 0-0.2 ppm. For See
 454 the SI, Section S4 for GHGSat images using the standard 0-2 ppm scale applied for most retrieval images in this
 455 study. Cloud-free surface imagery © 2023 Google Earth, CNES/Airbus, Maxar Technologies, USDA/FPAC/GEO.
 456

457 Sentinel-2 imagery is significantly noisier than most other tested satellites. The November 18th
 458 Kayrros retrieval in Figure 6 shows noticeable enhancements, comparable in intensity to the true
 459 emission, along the water feature and the highway, as well as northwest of the release site. In
 460 such a noise environment, knowledge of the emission location and access to images known not
 461 to contain emissions, such as panel f) may assist in correct identification of the true emission.
 462 See the SI, Section S4 for all masked and unmasked retrieval images from all satellites.
 463

464 Maxar correctly detected emissions as small as 0.0332 [0.0328, 0.0336] t/h using their
 465 WorldView-3 satellite on October 29th. Interestingly, their retrieval algorithm does not appear to
 466 introduce high-concentration artifacts over the water body (although that is not the case for all

467 teams analyzing WorldView-3 data, as shown in the SI, Section S4). The full retrieval image for
468 the October 29th retrieval shows concentration enhancement artifacts of comparable magnitude to
469 the correctly-detected emission at several points in the image. However, these artifacts are
470 largely conformal with surface features visible in optical imagery.

471
472 In addition to the known location, Maxar tasked its WorldView-3 satellite without notice to the
473 Stanford team on November 24th, a holiday in the United States. These data were shared with all
474 teams, but Maxar did not submit an unmasked image for the November 24th retrieval, although
475 they did for the zero-emission November 5th retrieval, shown in panel i). As a result, Maxar and
476 all other participating teams were able to compare satellite data from active testing days with
477 data that they knew very likely did not contain methane enhancements. As a result, these teams
478 had information in addition to the known release location that would not necessarily be available
479 in the field. As a result, we cannot definitively conclude from this study whether Maxar or other
480 teams would successfully identify emissions as small as 0.0332 [0.0328, 0.0336] t/h in the field.
481 Future testing, likely with multiple potential source locations, is needed to more rigorously
482 assess field-realistic detection limits of all satellites tested in this study.

483

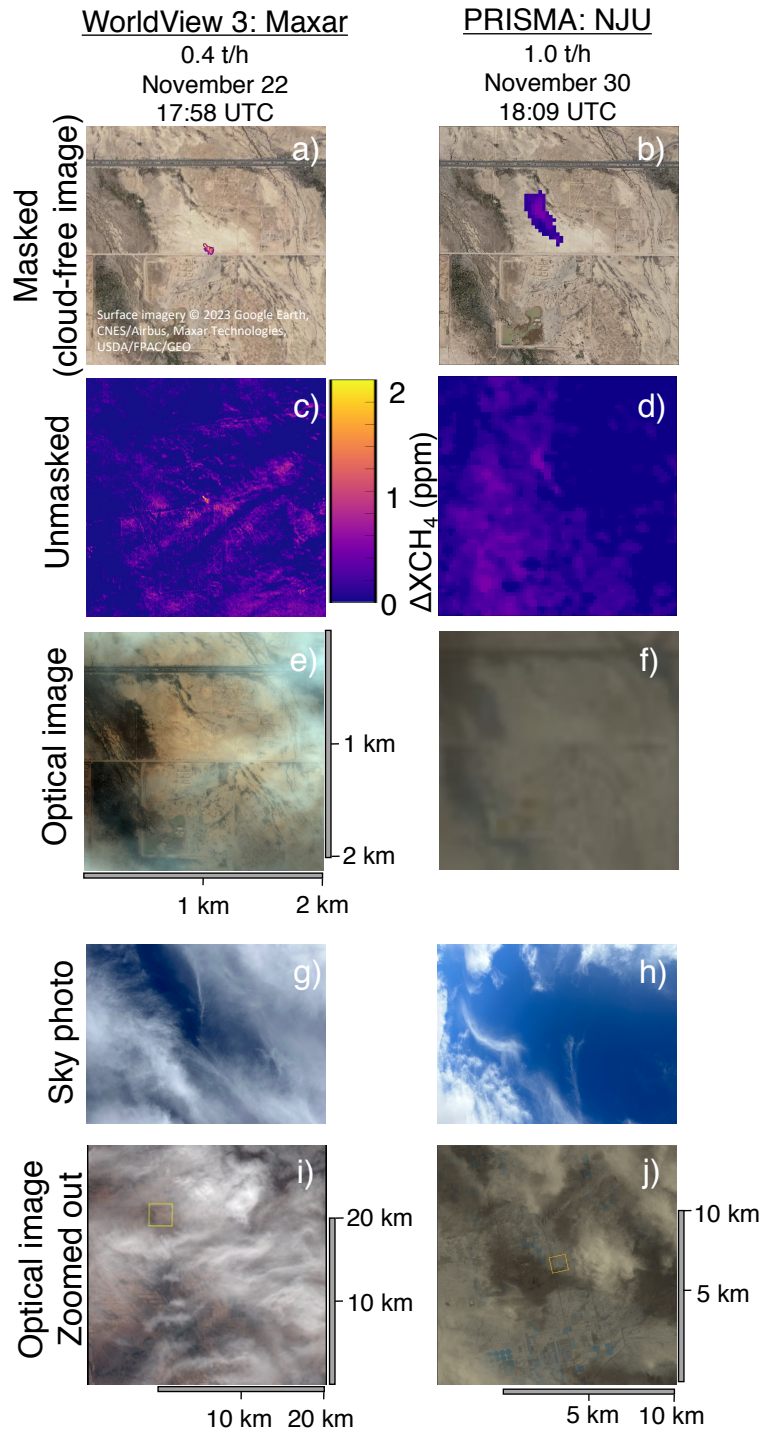
484 **2.4 The role of clouds**

485 Because water vapor is highly absorptive in the methane-active infrared frequencies targeted by
486 all nine methane-sensing satellites tested in this study, cloud cover can impede or prevent valid
487 satellite-based methane measurements. Although our Arizona test site was selected in part due to
488 its arid, relatively low-cloud climate, periodic cloud cover occurred to varying degrees
489 throughout the testing period.

490

491 The treatment of clouds varied across teams, with some filtering images due to cloud cover more
492 aggressively than others. LARS filtered the November 22nd WorldView-3 retrieval, shown in
493 Figure 7a, noting “the image is cloudy but we see some enhancement.” Kayrros and Maxar
494 correctly detected the 0.433 [0.430, 0.436] t/h emission for the same measurement, while NJU
495 reported a non-detection.

496



497
 498 Figure 7. Cloudy days with successful methane detections. a) and b) show masked methane emissions from
 499 WorldView-3 and PRISMA above a cloud-free standard background © 2023 Google Earth, CNES/Airbus, Maxar
 500 Technologies, USDA/FPAC/GEO. c) and d) show corresponding unmasked images. e) and f) show optical images
 501 of the same 2x2 km scene collected by each satellite. g) and h) show photographs of the sky, taken by Stanford
 502 researchers on smartphones at the time of each overpass. i) and j) show zoomed-out versions of the optical images
 503 shown in e) and f), with different length scales than the other panels.

504 This highlights that accurately interpreting the results of field measurements from each of these
505 teams requires an understanding of both detection performance and data filtering processes as a
506 function of cloud cover.

507
508 Stanford researchers took photographs of the sky coincident with most satellite overpasses to
509 document cloud cover, shown in full in the SI, Section S4. The photograph for the November
510 22nd WorldView-3 overpass, Figure 7g, appears to show significant thick cloud cover. However,
511 analysis of optical WorldView-3 imagery from this measurement, Figure 7e, shows that the area
512 immediately above the test site was relatively cloud-free even though the broader area was
513 experiencing significant cloud cover, shown in Figure 7i.

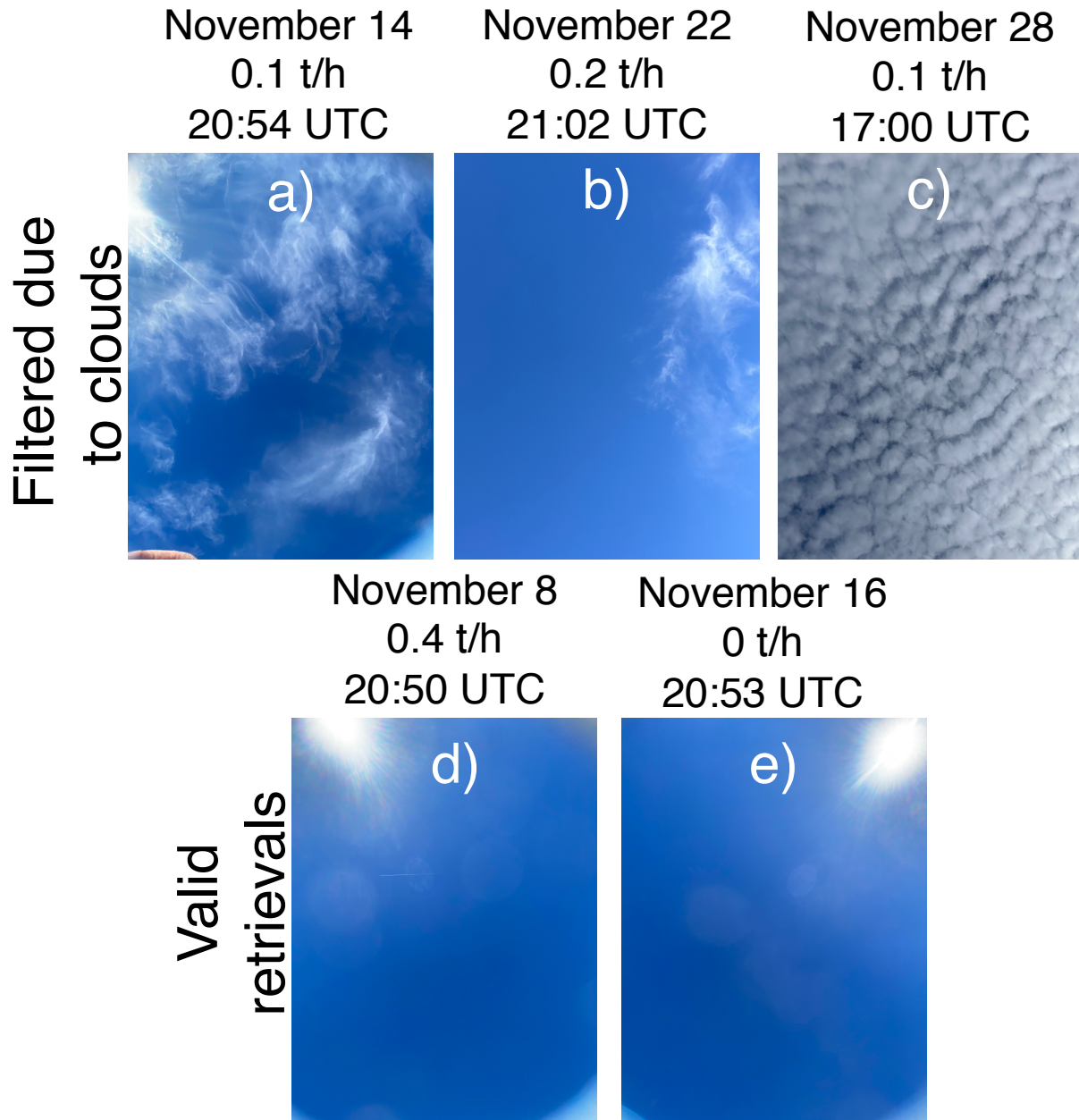
514
515 Analysis of the November 30th PRISMA measurement, shown in the second column of Figure 7
516 adds further nuance to the question of cloud cover. The sky photograph in Figure 7h shows the
517 presence of thin clouds. However, the optical image collected by PRISMA in Figure 7f shows no
518 clouds within the 2x2 km square surrounding the release site. The photographed clouds are only
519 visible in the larger, 14x14 km image in Figure 7j, which demonstrates that clouds are too far
520 away from the release site to interfere with the 0.98 [0.88, 1.08] t/h methane plume, which was
521 correctly detected by LARS, NJU, and Maxar.

522
523 These two cases demonstrate that only limited information regarding cloud cover can be
524 determined from single-frame sky photographs taken from the ground. This is particularly true
525 without clear orientation information, which is not available for the smartphone-based
526 photographs used in this study.

527
528 Figure 8 shows sky photographs of all dates with valid or operator-filtered GHGSat
529 measurements. Both days with valid measurements, one true positive and one true negative, were
530 essentially cloudless, as shown in Figure 8d-e. In addition, GHGSat filtered three retrievals due
531 to clouds. Of the three days filtered due to cloud cover, one was fully overcast (Figure 8c), while
532 two had thin clouds, shown in Figure 8a-b, also noted in the GHGSat report for those days. As
533 demonstrated above, it is difficult to determine from these sky photographs alone where these
534 clouds were in relation to the release site.

535

GHGSat-coincident sky photos



536
537 Figure 8. Ground-perspective sky photos for GHGSat-C measurements. a-c) correspond to measurements filtered
538 due to cloud cover. d) and e) correspond to valid retrievals, including one true positive detection and one true
539 negative non-detection. GHGSat-C satellites do not collect optical imagery, making it difficult to directly compare
540 ground-perspective photographs with satellite-perspective optical imagery.

541 GHGSat did not submit unmasked retrieval images for operator-filtered measurements (these
542 images were requested from all teams, but were not required as a condition of participation in
543 this test). Furthermore, GHGSat does not collect optical imagery in visible frequencies, so none

544 could be submitted. As a result, we can draw only limited conclusions about the role of cloud
545 cover in GHGSat’s ability to conduct valid measurements with the GHGSat-C satellite model.
546

547 Future satellite-focused controlled methane release tests should further investigate the role of
548 cloud cover. This should include conducting testing in cloudier locations. In addition, sky
549 photographs should be replaced by or supplemented with passively-collected time series of
550 panoramic, georeferenced sky time series, e.g. using a fisheye camera, e.g., as used in solar
551 forecasting systems (Sun et al., 2018). This, together with optical images collected by satellites
552 (when available), will allow a more systematic evaluation of the capabilities of the tested
553 systems as a function of cloud cover. Such analysis should include assessment of the effect of
554 clouds on detection sensitivity and quantification performance, as well as their role in preventing
555 collection of valid measurements. These cloud-informed performance findings will be
556 indispensable in regional analysis of satellite-based methane remote sensing data, including its
557 incorporation into emissions inventories.
558

559 **3 Discussion**

560 This work demonstrates that all tested satellites are capable of detecting and quantifying methane
561 emissions. All eight satellites given the opportunity detected methane emissions, with overall
562 quantification accuracy similar, in percent terms, to aircraft-based methane sensing systems. This
563 highlights the large suite of satellite-based tools available to detect and quantify methane point
564 sources across the globe.
565

566 Detection limits appear to improve with smaller swath width and pixel size, and with higher
567 spectral resolution. Global-coverage satellites such as LandSat 8/9 and Sentinel-2, with swaths of
568 185 and 290 km, respectively, and spectral resolution 20-650 times coarser than the
569 hyperspectral instruments (EnMAP, PRISMA, GF5, ZY1, HJ2B, and GHGSat), have higher
570 detection limits. See the SI, Section S2 for additional discussion of spectral resolution. Our
571 results are consistent with (Gorroño et al., 2023), whose simulation-based approach suggests that
572 such instruments have a best-case minimum detection limit of roughly 1 t/h. Targeted satellites
573 with swaths of 30-60 km, including EnMAP, GF5, PRISMA, and ZY1 (EnMAP, 2023; Liu et al.,
574 2019; OHBI, 2022; Song et al., 2022), all reliably saw emissions of ~1 t/h. Of these, only
575 PRISMA has had the opportunity to be tested with emission fluxes below 1 t/h, correctly
576 detecting 0.413 [0.410, 0.417] t/h, the smallest emission given to PRISMA. GHGSat correctly
577 detected 0.401 [0.399, 0.403] t/h, with quantification accuracy within $\pm 20\%$, using their
578 GHGSat-C-series satellite, with a swath width of 12 km. Estimates for smaller emission sizes
579 were filtered due to clouds, but in previous testing GHGSat successfully detected an 0.197
580 [0.187, 0.208] t/h emission and quantified it with similar accuracy, suggesting that the system
581 may be capable of seeing emissions even smaller than 200 kg/h.
582

583 Maxar successfully detected emissions as low as 0.0332 [0.0328, 0.0336] t/h using the
584 WorldView-3 satellite, with swath width 13.1 km. Two teams successfully detected emissions
585 below 0.1 t/h using WorldView-3, while two teams applied more conservative criteria and
586 detected only emissions above 0.5 t/h. Although Maxar has a coarser spectral resolution than
587 hyperspectral instruments, its very high spatial resolution enables heightened sensitivity.
588

589 In the high-emission New Mexico Permian basin oil and natural gas system, using 2019 emission
590 levels, a comprehensive measurement campaign with a constellation of satellites detecting all
591 emissions above 1 t/h would find 20% of emissions from oil and gas well sites, rising to 62% for
592 a satellite detecting emissions above 0.2 t/h, and 83% above 0.03 t/h (Sherwin et al., 2023b).
593 These fractions are upper-bound estimates both because near-real-time comprehensive coverage
594 would be challenging for satellite systems and because the underlying emission size distribution
595 estimate may be conservative for emissions below roughly 50 kg/hr (Sherwin et al., 2023b). In
596 lower-emitting basins such as the Denver-Julesburg, each of these systems would see a much
597 smaller fraction of total emissions, highlighting the need for a variety of technology approaches,
598 tailored to regional system characteristics (Sherwin et al., 2023b).
599

600 Note that the detection results presented in this paper reflect system performance with a known
601 source location under favorable desert climate conditions. These results may not translate to field
602 performance in different environments and with less foreknowledge about the location of
603 possible sources.
604

605 Unmasked methane retrieval fields, submitted by all teams, suggest that achievable detection
606 limits may be higher in practice for some satellites. In some cases, these images contain
607 background artifacts with estimated methane enhancements comparable in magnitude and
608 qualitatively similar in shape to the detected methane plumes. However, in many of these
609 retrieval fields, particularly for larger emissions, the true methane plume is unambiguous. It is
610 noteworthy that some teams correctly flagged likely background artifacts in blinded submissions,
611 but such georeferenced quality flagging was not required of all participating teams, although
612 doing so may be advisable in future tests.
613

614 The role of surface features, such as water bodies, in creating apparent methane enhancements
615 should be explored further. For example, the retrieval field for the 0.401 [0.399, 0.403] t/h
616 GHGSat measurement shows an apparent methane enhancement over a water body that is similar
617 in magnitude to the detected plume. However, if this is a known characteristic of the algorithm,
618 then such artifacts could be automatically or manually filtered out, leaving only the clear
619 methane plume at the release site. The water body appears as a flagged region in all data reported
620 by GHGSat, indicating that their system is capable of identifying potential confounding factors
621 such as water bodies and differentiating any resulting artifacts from true methane emissions.
622

623 Clouds add several levels of complexity to satellite-based methane sensing. The water vapor in
624 clouds interferes with the frequencies all tested satellites use to identify methane enhancements.
625 Heavy cloud cover essentially prevents valid satellite-based methane sensing. This test
626 demonstrates that it is possible in some circumstances to detect and quantify methane emissions
627 even in the presence of nearby patchy or thin clouds. However, it is unclear in some cases
628 whether these detected emissions would have been distinguishable from background noise, e.g.
629 artifacts caused by clouds or highly reflective/absorptive surface features, in the absence of a
630 known source location and reasonable anticipation of the presence of an emission due to an
631 ongoing test.
632

633 Different teams employed different filtering criteria. GHGSat excluded all GHGSat-C
634 measurements with cloud cover. Maxar and Kayrros used WorldView-3 to successfully detected

635 a 0.433 [0.430, 0.436] t/h emission on a cloudy day on November 22nd, while LARS filtered the
636 measurement due to clouds and NJU reported a non-detection.

637
638 Future testing should characterize the cloud conditions under which valid point-source methane
639 measurements can and cannot be conducted with each satellite-based system. In addition, future
640 work should characterize the effect of partial cloud cover on detection and quantification
641 performance. Understanding these two factors will be critical when interpreting the results of
642 large-scale satellite-based methane measurement campaigns, which will inevitably encounter
643 interference from clouds. Cloud cover varies widely across oil and gas-producing regions, with
644 limited clouds in arid areas such as the Permian basin in Texas and New Mexico, and significant
645 cloud cover in more temperate producing regions such as the Appalachian basin in the eastern
646 United States and the Williston basin in the midwestern United States (NASA, 2023).

647
648 It is noteworthy that even under cloud-free conditions, a targeted satellite overpass is not
649 guaranteed to produce valid data. Errors in tasking software, as well as onboard hardware upsets
650 can prevent valid data collection. The incidence of both in this paper may not be representative
651 of field performance for the tested technologies. Additional data collection, ideally from field
652 data, would be needed to accurately quantify the incidence of data collection failure, and further
653 location-specific analysis of cloud trends would be needed to understand the impact of cloud
654 cover on satellite data collection capabilities in a specific area.

655
656 Wind speed remains a major driver of uncertainty in satellite-based methane point source
657 quantification. Moving from wind reanalysis data to in situ wind measurements substantially
658 reduces scatter around the line of best fit, as was also the case in other work from the same group
659 (Sherwin et al., 2023b). In addition, in situ wind measurements show considerable temporal
660 variability in wind speed and direction over the multi-minute timescales most relevant to plume
661 formation.

662
663 In the field, winds are generally only available from reanalysis data, which capture temporal,
664 spatial, and directional variability with much lower fidelity than on-the-ground wind
665 measurements. Advances in the spatial and temporal fidelity of wind reanalysis products, as well
666 as their accuracy, could help improve methane remote sensing. In addition, it may be possible to
667 entirely eliminate reliance on wind speed, e.g. by inferring emission rate information solely from
668 plume shapes as in reference (Jongaramrungruang et al., 2022).

669
670 It is important to note that conducting this test did require the release of considerable amounts of
671 methane into the atmosphere. We estimate total emissions from the satellite testing discussed in
672 this paper at 7.7 t(CH₄), discussed further in the SI, Section S1.5. However, this pales in
673 comparison with anthropogenic emissions occurring across the globe. Lauvaux et al. identify
674 over 1000 emission sources across the world emitting at least 7.7 t(CH₄) every hour, in some
675 cases over 50 times as much every hour (Lauvaux et al., 2022). If this work assists in
676 accelerating mitigation of even one of these emissions by even a single hour, e.g. by ensuring
677 key decision-makers view satellite-based methane detection and quantification as reliable, we
678 will have broken even from a methane emissions perspective.

679
680

681 The findings presented here demonstrate that at least eight distinct satellite systems from three
682 continents are capable of detecting methane point sources of 1.5 t/h or less. Furthermore, this
683 study more systematically probes lower detection limits of these systems, two teams detecting
684 emissions below 0.1 t/h, the first time to our knowledge that such performance has been
685 demonstrated in a single-blind test of satellite-based methane sensing systems.
686

687 These satellites can play an important role in reducing methane emissions through existing
688 regulatory pathways, both in the United States and internationally. The US Environmental
689 Protection Agency’s proposed update to rules governing methane emissions from oil and natural
690 gas production includes a super-emitter response program, in which approved third-party data
691 providers can flag identified emissions above 0.1 t/h, obliging operators to investigate further
692 and, if necessary, take action to halt any further emissions (EPA, 2022). A proposed update to
693 the EPA Greenhouse Gas Reporting Program also includes a new category of “Other large
694 release” for inclusion in company emissions reports (EPA, 2023). The Methane Alert and
695 Response Systems, part of the United Nations’ International Methane Emissions Observatory,
696 uses vetted satellite data to notify governments, and in some cases operators, of large emissions
697 detected by satellite, with the aim of mitigating these emissions (IMEO, 2023). The eight
698 satellite systems tested with at least one nonzero emission in this study can provide high-quality
699 data to each of these programs.
700

701 In coming years, the Carbon Mapper and MethaneSAT systems will launch, alongside additional
702 satellites in some of the constellations tested here (Jacob et al., 2022). The airplane-mounted
703 precursors to both the Carbon Mapper and MethaneSAT systems have conducted substantial
704 single-blind testing of their point-source detection and quantification capabilities (Rutherford et
705 al., 2023; Chulakadabba et al., 2023; El Abbadi et al., 2023), but the satellites will require
706 additional tests. Furthermore, the NASA Earth Surface Mineral Dust Source Investigation
707 (EMIT) system, which launched shortly before our testing began (Wang and Lee, 2022), has
708 already reported detecting methane emissions in the field and should be tested, along with the
709 HJ2 system, in future single-blind controlled methane releases.
710

711 The tools exist for multi-lateral global methane monitoring efforts, with satellites from multiple
712 countries and continents able to independently assess emissions from regions of interest. The
713 single-blind test conducted here is a step toward ensuring that stakeholders across the world have
714 confidence in the methane emissions these satellite systems find at oil and gas facilities, landfills,
715 coal mines, and other emitting infrastructure. This will help satellites achieve their potential to
716 not only detect and quantify large methane emissions, but to inspire meaningful action to reduce
717 emissions of this powerful greenhouse gas.
718

719 **4 Data and code availability**

720 All data and code required to reproduce the figures and analysis in this paper are available at:
721 <https://github.com/sahar-elabbadi/SU-Controlled-Releases-2022>. Underlying spectral imagery
722 will not be made directly available through this study, but for many satellites tested in this study
723 these spectral data can be acquired either for free or for purchase for via platforms discussed in
724 the SI, Section S2.

725
726

5 Abbreviations

ADED	Advancing Development of Emissions Detection
ASI	Italian Space Agency
CNG	Compressed Natural Gas
EMIT	Earth Surface Mineral Dust Source Investigation
EnMAP	Environmental Mapping and Analysis Program
GF5	Gaofen 5
GSC	GHGSat-C (satellite)
HJ2	Huanjing 2
IME	Integrated Mass Enhancement
kg/h	Kilograms per hour
LARS	Land and Atmosphere Remote Sensing
LS	LandSat
METEC	Methane Emissions Technology Evaluation Center
NASA	National Aeronautics and Space Administration
NJU	Nanjing University
NOAA	National Oceanographic and Atmospheric Administration
OHB	Orbitale Hochtechnologie Bremen
PRISMA	PRecursore IperSpettrale della Missione Applicativa
UPV	Universitat Politècnica de València
USGS	United States Geological Survey
SRON	Stichting Ruimte Onderzoek Nederland
SWIR	Short-wave Infrared
TROPOMI	TROPOspheric Monitoring Instrument
t/h	Metric tons per hour
VNIR	Visible to Near Infrared
WAV-P	Wide-Angle Fabry-Perot
WV3	WorldView-3
ZY1	Ziyuan 1

727

6 Acknowledgments

729 We acknowledge C. de Franchis, C. Giron, and A. Groshenry of Kayrros; and Z. Mouton, W.
730 Kingwill, and R. Huppertz from Orbio Earth for their participation in this test. A. Esparza, L.
731 Clark-Squire, J.F. Gauthier, M. Girard, D. Jervis, R. Mattson, J. McKeever, A. Newhook, and M.
732 Turenne of GHGSat; J. Gorroño Viñegla, J. Roger Juan, and L. Guanter Palomar of LARS; Chen
733 H., Li F., and Zhang H. of NJU; A. Hayden, J. Jonik, and J. Christy of Maxar both for
734 participating in the test and for coordinating tasking and data sharing from key satellites. We
735 acknowledge the German Aerospace Center and the Italian Space Agency for tasking the
736 EnMAP and PRISMA satellites, respectively. Rawhide Leasing (Dana Walker) and Volta
737 Fabrication (Mike Brandon, Walt Godsil, S.M., Merritt Norton) provided indispensable
738 operational, logistical, and planning support for the experiment.

739 **7 Author contributions**

740 Conceptualization – EDS and ARB. Methods – EDS, SHE, YC, JSR, and ARB. Software – EDS,
741 PMB, ZC, and SHE. Validation – EDS. Formal analysis – EDS, PMB, ZZ, and SHE.
742 Investigation – EDS and SHE. Resources – ARB. Data curation – EDS, PMB, YC, ZZ, ZC, JSR,
743 and SHE. Writing: original draft – EDS. Writing: review and editing – All authors. Supervision –
744 EDS and ARB. Project administration – EDS, SEA, and ARB. Funding acquisition – EDS, SHE,
745 ARB. This work was supported by the US Department of Energy, Office of Science through
746 contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US
747 Department of Energy. The US government retains, and the publisher, by accepting the article
748 for publication, acknowledges that the US government retains, a non-exclusive, paid-up,
749 irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or
750 allow others to do so, for US government purposes.

751 **8 Supplementary information available**

752 The online version contains supplementary material.

753 **9 Funding sources**

754 This study was funded by: The Environmental Defense Fund, the Global Methane Hub, the
755 International Methane Emissions Observatory, and the Stanford Natural Gas Initiative, an
756 industry consortium that supports independent research at Stanford University.

757 **10 Competing interests**

758 ARB serves on the Science and Measurement advisory committee of Carbon Mapper, which
759 plans to launch a methane-sensing satellite. YC and ZZ previously worked as interns at Carbon
760 Mapper. The remaining authors have no competing interests to declare.

761 **11 References**

762 Bell, C., Rutherford, J., Brandt, A., Sherwin, E., Vaughn, T., and Zimmerle, D.: Single-blind
763 determination of methane detection limits and quantification accuracy using aircraft-based
764 LiDAR, *Elementa: Science of the Anthropocene*, 10, 00080,
765 <https://doi.org/10.1525/elementa.2022.00080>, 2022.

766 Bell, C., Ilonze, C., Duggan, A., and Zimmerle, D.: Performance of Continuous Emission
767 Monitoring Solutions under a Single-Blind Controlled Testing Protocol, *Environ. Sci. Technol.*,
768 57, 5794–5805, <https://doi.org/10.1021/acs.est.2c09235>, 2023.

769 Bell, C. S., Vaughn, T. L., Zimmerle, D., Herndon, S. C., Yacovitch, T. I., Heath, G. A., Pétron,
770 G., Edie, R., Field, R. A., Murphy, S. M., Robertson, A. M., and Soltis, J.: Comparison of
771 methane emission estimates from multiple measurement techniques at natural gas production
772 pads, *Elementa: Science of the Anthropocene*, 5, 79, <https://doi.org/10.1525/elementa.266>, 2017.

773 Bell, C. S., Vaughn, T., and Zimmerle, D.: Evaluation of next generation emission measurement
774 technologies under repeatable test protocols, *Elementa: Science of the Anthropocene*, 8, 32,
775 <https://doi.org/10.1525/elementa.426>, 2020.

776 Chen, Y., Sherwin, E. D., Berman, E. S. F., Jones, B. B., Gordon, M. P., Wetherley, E. B., Kort,
777 E. A., and Brandt, A. R.: Quantifying Regional Methane Emissions in the New Mexico Permian
778 Basin with a Comprehensive Aerial Survey, *Environ. Sci. Technol.*, 56, 4317–4323,
779 <https://doi.org/10.1021/acs.est.1c06458>, 2022.

780 Chulakadabba, A., Sargent, M., Lauvaux, T., Benmergui, J. S., Franklin, E., Miller, C. C.,
781 Wilzewski, J. S., Roche, S., Conway, E., Sour, A. H., Sun, K., Luo, B., Hawthorne, J., Samra, J.,
782 Daube, B. C., Liu, X., Chance, K., Li, Y., Gautam, R., Omara, M., Rutherford, J. S., Sherwin, D.,
783 Brandt, A., and Wofsy, S. C.: Methane Point Source Quantification Using MethaneAIR: A New
784 Airborne Imaging Spectrometer, 2023.

785 Cusworth, D. H., Duren, R. M., Thorpe, A. K., Olson-Duvall, W., Heckler, J., Chapman, J. W.,
786 Eastwood, M. L., Helmlinger, M. C., Green, R. O., Asner, G. P., Dennison, P. E., and Miller, C.
787 E.: Intermittency of Large Methane Emitters in the Permian Basin, *Environ. Sci. Technol. Lett.*,
788 8, 567–573, <https://doi.org/10.1021/acs.estlett.1c00173>, 2021.

789 Cusworth, D. H., Thorpe, A. K., Ayasse, A. K., Stepp, D., Heckler, J., Asner, G. P., Miller, C. E.,
790 Yadav, V., Chapman, J. W., Eastwood, M. L., Green, R. O., Hmiel, B., Lyon, D. R., and Duren,
791 R. M.: Strong methane point sources contribute a disproportionate fraction of total emissions
792 across multiple basins in the United States, *Proc. Natl. Acad. Sci. U.S.A.*, 119, e2202338119,
793 <https://doi.org/10.1073/pnas.2202338119>, 2022.

794 Duren, R. M., Thorpe, A. K., Foster, K. T., Rafiq, T., Hopkins, F. M., Yadav, V., Bue, B. D.,
795 Thompson, D. R., Conley, S., Colombi, N. K., Frankenberg, C., McCubbin, I. B., Eastwood, M.
796 L., Falk, M., Herner, J. D., Croes, B. E., Green, R. O., and Miller, C. E.: California’s methane
797 super-emitters, *Nature*, 575, 180–184, <https://doi.org/10.1038/s41586-019-1720-3>, 2019.

798 El Abbadi, S., Chen, Z., Burdeau, P., Rutherford, J., Chen, Y., Zhang, Z., Sherwin, E., and
799 Brandt, A.: Comprehensive evaluation of aircraft-based methane sensing for greenhouse gas
800 mitigation, *Engineering*, <https://doi.org/10.31223/X51D4C>, 2023.

801 EnMAP: EnMAP: Mission, EnMAP, EnMAP, 2023.

802 EPA: Standards of Performance for New, Reconstructed, and Modified Sources and Emissions
803 Guidelines for Existing Sources: Oil and Natural Gas Sector Climate Review, 40 CFR Part 60,
804 87, 2022.

805 EPA: Greenhouse Gas Reporting Rule: Revisions and Confidentiality Determinations for
806 Petroleum and Natural Gas Systems, 40 CFR Part 98, 88, 2023.

807 ESA: eoPortal: PRISMA (Hyperspectral), European Space Agency, Paris, France, 2012.

808 ESA: Sentinel-2, European Space Agency, Paris, France, 2021a.

809 ESA: Sentinel-5P, European Space Agency, Paris, France, 2021b.

810 ESA: About GHGSat, European Space Agency, Paris, France, 2022a.

811 ESA: Earth Online: Worldview-3, European Space Agency, Paris, France, 2022b.

812 ESA: Sentinel Online: Sentinel-2, European Space Agency, Paris, France, 2022c.

813 GHGSat: Global leader in remote sensing of greenhouse gas, GHGSat, Montreal, Canada, 2022.

814 Gorroño, J., Varon, D. J., Irakulis-Loitxate, I., and Guanter, L.: Understanding the potential of
815 Sentinel-2 for monitoring methane point emissions, *Atmos. Meas. Tech.*, 16, 89–107,
816 <https://doi.org/10.5194/amt-16-89-2023>, 2023.

817 Guanter, L., Irakulis-Loitxate, I., Gorroño, J., Sánchez-García, E., Cusworth, D. H., Varon, D. J.,
818 Cogliati, S., and Colombo, R.: Mapping methane point emissions with the PRISMA spaceborne
819 imaging spectrometer, *Remote Sensing of Environment*, 265, 112671,
820 <https://doi.org/10.1016/j.rse.2021.112671>, 2021.

821 Hayden, A. and Christy, J.: Maxar’s WorldView-3 Enables Low-Concentration Methane
822 Detection from Space, <https://doi.org/10.31223/X51T1C>, 15 June 2023.

823 IMEO: Methane Alert and Response System (MARS), United Nations Environment Programme,
824 International Methane Emissions Observatory, Paris, France, 2023.

825 Irakulis-Loitxate, I., Guanter, L., Liu, Y.-N., Varon, D. J., Maasackers, J. D., Zhang, Y.,
826 Chulakadabba, A., Wofsy, S. C., Thorpe, A. K., Duren, R. M., Frankenberg, C., Lyon, D. R.,
827 Hmiel, B., Cusworth, D. H., Zhang, Y., Segl, K., Gorroño, J., Sánchez-García, E., Sulprizio, M.
828 P., Cao, K., Zhu, H., Liang, J., Li, X., Aben, I., and Jacob, D. J.: Satellite-based survey of
829 extreme methane emissions in the Permian basin, *Sci. Adv.*, 7, eabf4507,
830 <https://doi.org/10.1126/sciadv.abf4507>, 2021.

831 Irakulis-Loitxate, I., Gorroño, J., Zavala-Araiza, D., and Guanter, L.: Satellites Detect a Methane
832 Ultra-emission Event from an Offshore Platform in the Gulf of Mexico, *Environ. Sci. Technol.*
833 *Let.*, 9, 520–525, <https://doi.org/10.1021/acs.estlett.2c00225>, 2022a.

834 Irakulis-Loitxate, I., Guanter, L., Maasackers, J. D., Zavala-Araiza, D., and Aben, I.: Satellites
835 Detect Abatable Super-Emissions in One of the World’s Largest Methane Hotspot Regions,
836 *Environ. Sci. Technol.*, 56, 2143–2152, <https://doi.org/10.1021/acs.est.1c04873>, 2022b.

837 Jacob, D. J., Varon, D. J., Cusworth, D. H., Dennison, P. E., Frankenberg, C., Gautam, R.,
838 Guanter, L., Kelley, J., McKeever, J., Ott, L. E., Poulter, B., Qu, Z., Thorpe, A. K., Worden, J.
839 R., and Duren, R. M.: Quantifying methane emissions from the global scale down to point
840 sources using satellite observations of atmospheric methane, *Atmos. Chem. Phys.*, 22, 9617–
841 9646, <https://doi.org/10.5194/acp-22-9617-2022>, 2022.

842 Jervis, D., McKeever, J., Durak, B. O. A., Sloan, J. J., Gains, D., Varon, D. J., Ramier, A.,
843 Strupler, M., and Tarrant, E.: The GHGSat-D imaging spectrometer, *Atmos. Meas. Tech.*, 14,
844 2127–2140, <https://doi.org/10.5194/amt-14-2127-2021>, 2021.

845 Jia, M., Li, F., Zhang, Y., Wu, M., Li, Y., Feng, S., Wang, H., Chen, H., Ju, W., Lin, J., Cai, J.,
846 Zhang, Y., and Jiang, F.: The Nord Stream pipeline gas leaks released approximately 220,000

847 tonnes of methane into the atmosphere, *Environmental Science and Ecotechnology*, 12, 100210,
848 <https://doi.org/10.1016/j.es.2022.100210>, 2022.

849 Jongaramrungruang, S., Thorpe, A. K., Matheou, G., and Frankenberg, C.: MethaNet – An AI-
850 driven approach to quantifying methane point-source emission from high-resolution 2-D plume
851 imagery, *Remote Sensing of Environment*, 269, 112809,
852 <https://doi.org/10.1016/j.rse.2021.112809>, 2022.

853 Kayrros: A partner for today and the future, agile with technology and with a smarter approach
854 to data, Kayrros, Paris, France, 2022.

855 Lauvaux, T., Giron, C., Mazzolini, M., d'Aspremont, A., Duren, R., Cusworth, D., Shindell, D.,
856 and Ciaia, P.: Global Assessment of Oil and Gas Methane Ultra-Emitters, *Science*, 375, 557–
857 561, <https://doi.org/10.31223/X5NS54>, 2022.

858 Liu, Y.-N., Zhang, J., Zhang, Y., Sun, W.-W., Jiao, L.-L., Sun, D.-X., Hu, X.-N., Ye, X., Li, Y.-
859 D., Liu, S.-F., Cao, K.-Q., Chai, M.-Y., and Zhou, W.-Y.-N.: The Advanced Hyperspectral
860 Imager: Aboard China's GaoFen-5 Satellite, *IEEE Geosci. Remote Sens. Mag.*, 7, 23–32,
861 <https://doi.org/10.1109/MGRS.2019.2927687>, 2019.

862 Luo, H., Li, Z., Wu, Y., Qiu, Z., Shi, H., Wang, Q., and Xiong, W.: Greenhouse Gases
863 Monitoring Instrument on GaoFen-5 Satellite-II: Optical Design and Evaluation, *Remote
864 Sensing*, 15, 1105, <https://doi.org/10.3390/rs15041105>, 2023.

865 NASA: NASA Earth Observations Cloud Fraction (1 Month TERRA/MODIS), National
866 Aeronautics and Space Administration, Washington, D.C., 2023.

867 OHBI: Satellites & Missions: PRISMA, Orbitale Hochttechnologie Bremen Italia S.p.A., Milan,
868 Italy, 2022.

869 Orbio: Actionable Methane Intelligence: Filling the global methane gap with asset-level
870 emissions data, Orbio, Köln, Germany, 2023.

871 Pandey, S., Gautam, R., Houweling, S., van der Gon, H. D., Sadavarte, P., Borsdorff, T.,
872 Hasekamp, O., Landgraf, J., Tol, P., van Kempen, T., Hoogeveen, R., van Hees, R., Hamburg, S.
873 P., Maasackers, J. D., and Aben, I.: Satellite observations reveal extreme methane leakage from a
874 natural gas well blowout, *Proc. Natl. Acad. Sci. U.S.A.*, 116, 26376–26381,
875 <https://doi.org/10.1073/pnas.1908712116>, 2019.

876 Ravikumar, A. P., Sreedhara, S., Wang, J., Englander, J., Roda-Stuart, D., Bell, C., Zimmerle,
877 D., Lyon, D., Mogstad, I., Ratner, B., and Brandt, A. R.: Single-blind Inter-comparison of
878 Methane Detection Technologies - Results from the Stanford/EDF Mobile Monitoring
879 Challenge, *Elementa: Science of the Anthropocene*, 7, 29, <https://doi.org/10.1525/elementa.373>,
880 2019.

881 Roger, J., Irakulis-Loitxate, I., Valverde, A., Gorroño, J., Chabrilat, S., Brell, M., and Guanter,
882 L.: High-resolution methane mapping with the EnMAP satellite imaging spectroscopy mission,
883 *Atmospheric Sciences*, <https://doi.org/10.31223/X5M65Z>, 2023.

884 Rutherford, J., Sherwin, E., Chen, Y., Aminfard, S., and Brandt, A.: Evaluating methane
885 emission quantification performance and uncertainty of aerial technologies via high-volume
886 single-blind controlled releases, *Oil, Gas, and Energy*, <https://doi.org/10.31223/X5KQ0X>, 2023.

887 Sánchez-García, E., Gorroño, J., Irakulis-Loitxate, I., Varon, D. J., and Guanter, L.: Mapping
888 methane plumes at very high spatial resolution with the WorldView-3 satellite, *Environmental
889 Science & Technology*, 56, 10517–10529, <https://doi.org/10.1021/acs.est.1c08575>, 2022.

890 Scott, W.: Mapping Methane Emissions Using Maxar’s WorldView-3 Satellite, Maxar, Ann
891 Arbor, MI, USA, 2022.

892 Sherwin, E., Rutherford, J., Zhang, Z., Chen, Y., Wetherley, E., Yakovlev, P., Berman, E., Jones,
893 B., Thorpe, A., Ayasse, A., Duren, R., Brandt, A., and Cusworth, D.: Quantifying oil and natural
894 gas system emissions using one million aerial site measurements, In Review,
895 <https://doi.org/10.21203/rs.3.rs-2406848/v1>, 2023a.

896 Sherwin, E. D., Chen, Y., Ravikumar, A. P., and Brandt, A. R.: Single-blind test of airplane-
897 based hyperspectral methane detection via controlled releases, *Elementa: Science of the
898 Anthropocene*, 9, 00063, <https://doi.org/10.1525/elementa.2021.00063>, 2021.

899 Sherwin, E. D., Rutherford, J. S., Chen, Y., Aminfard, S., Kort, E. A., Jackson, R. B., and
900 Brandt, A. R.: Single-blind validation of space-based point-source detection and quantification of
901 onshore methane emissions, *Sci Rep*, 13, 3836, <https://doi.org/10.1038/s41598-023-30761-2>,
902 2023b.

903 Song, Q., Ma, C., Liu, J., and Wei, H.: Quantifying ocean surface green tides using high-spatial
904 resolution thermal images, *Opt. Express*, 30, 36592, <https://doi.org/10.1364/OE.472479>, 2022.

905 Sun, Y., Szücs, G., and Brandt, A. R.: Solar PV output prediction from video streams using
906 convolutional neural networks, *Energy Environ. Sci.*, 8, 2018.

907 USGS: Landsat 8, United States Geological Survey, Washington, D.C., 2022.

908 Varon, D. J., Jacob, D. J., McKeever, J., Jervis, D., Durak, B. O. A., Xia, Y., and Huang, Y.:
909 Quantifying methane point sources from fine-scale satellite observations of atmospheric methane
910 plumes, *Atmos. Meas. Tech.*, 11, 5673–5686, <https://doi.org/10.5194/amt-11-5673-2018>, 2018.

911 Varon, D. J., McKeever, J., Jervis, D., Maasackers, J. D., Pandey, S., Houweling, S., Aben, I.,
912 Scarpelli, T., and Jacob, D. J.: Satellite Discovery of Anomalously Large Methane Point Sources
913 From Oil/Gas Production, *Geophys. Res. Lett.*, 46, 13507–13516,
914 <https://doi.org/10.1029/2019GL083798>, 2019.

915 Varon, D. J., Jervis, D., McKeever, J., Spence, I., Gains, D., and Jacob, D. J.: High-frequency
916 monitoring of anomalous methane point sources with multispectral Sentinel-2 satellite
917 observations, *Atmos. Meas. Tech.*, 14, 2771–2785, <https://doi.org/10.5194/amt-14-2771-2021>,
918 2021.

- 919 Wang, A. and Lee, J. J.: Methane ‘Super-Emitters’ Mapped by NASA’s New Earth Space
920 Mission, National Aeronautics and Space Administration, Pasadena, California, USA, 2022.
- 921 Xinhua: China launches new remote sensing satellite, XinhuaNet, 9th December, 2022.
- 922 Zhang, B., Guo, B., Zou, B., Wei, W., Lei, Y., and Li, T.: Retrieving soil heavy metals
923 concentrations based on GaoFen-5 hyperspectral satellite image at an opencast coal mine, Inner
924 Mongolia, China, Environmental Pollution, 300, 118981,
925 <https://doi.org/10.1016/j.envpol.2022.118981>, 2022.
- 926 Zhong, B., Yang, A., Liu, Q., Wu, S., Shan, X., Mu, X., Hu, L., and Wu, J.: Analysis Ready Data
927 of the Chinese GaoFen Satellite Data, Remote Sensing, 13, 1709,
928 <https://doi.org/10.3390/rs13091709>, 2021.
- 929 Zimmerle, D.: METEC Controlled Test Protocol: Survey Emission Detection And
930 Quantification, Colorado State University, Fort Collins, CO, USA, 2022.
- 931

