Response to Referee 1

Manuscript: Reviews and syntheses: Bridging Limnology and Satellite Methane Observations for Open Water Surfaces of Small Water Bodies (egusphere-2026-3707)

We are grateful to both referees and to the community commenter for reviews that have improved this manuscript considerably. Every comment has been addressed, and several prompted changes going beyond what was asked. Referee comments are reproduced in italics, followed by our response. Page and line numbers refer to the revised manuscript, except where a deletion is described, in which case they refer to the preprint, since the text exists only there. Referee comments quote the table numbers of the submitted version; because a duplicated table has been removed in revision, the former Tables 7 and 8 are now Tables 6 and 7, and our responses use the revised numbering.

The structure and writing style of the article is good. I could easily follow and read it, and the sections are well-differentiated. The topic is also very interesting, as I agree that emission from lakes and ponds are underestimated by current inventories.

We thank the referee for this generous assessment of the manuscript's structure and readability, and for the encouraging comment on the topic's importance.

There are some general pieces of information or values, like the emission rates of small ponds, sensitivity thresholds and spatial resolution from the satellite instruments, that are repeated too many times throughout the manuscript. I would suggest reducing the number of times the information comes up, as it gets repetitive.

We thank the referee for this observation, which prompted a systematic pass rather than isolated edits. We compiled every occurrence of the recurring quantities and reconciled them. The repeated passages removed in consequence are itemised under the specific comments below.

On a negative note, I am a bit concerned with the scope and base of the study. Without wanting to discourage or offend the authors, the study feels like a literature review at the beginning of a project, not a synthesis of a research area.

We thank the referee for raising this directly; it is a fair observation, and we can account for part of what produced it. The study was conceived, designed and written as a systematic literature review, following PRISMA screening throughout. Biogeosciences classifies submissions of this kind under the Reviews and syntheses manuscript type, and the title prefix was added at the request of the editorial office rather than chosen by us.

We are grateful to the referee for this observation, and we accept that carrying the Reviews and syntheses label creates a legitimate expectation of synthesis which the submitted version did not fully meet. Rather than rest on the question of classification, we have strengthened the manuscript with analysis it previously lacked. We now report, for each class of sensor, the minimum water body area that can be resolved and the corresponding fraction of the German inventory that falls within reach (Sect. 3.4.2, p. 15–16, lines 359–366, and Table 7, p. 16); the methane-specific instruments reach none of it, a result that could not be obtained from sensor specifications alone and required the inventory itself. We have added a treatment of each reviewed method setting out what further work would achieve and what constrains it, including a recommendation against one approach (Sect. 4.1, p. 18-19, lines 431–465). We now derive the detection limit of an area flux mapper from pixel dimension, wind speed and revisit frequency (Sect. 3.3.2, p. 13, lines 319–323), following the two studies the referee kindly

suggested, which yields a consequence visible only when the remote sensing and limnological literatures are read together: the temporal averaging that lowers the detection limit also removes the time resolution needed to resolve episodic ebullition. We hope the referee will find that these additions bring the paper closer to the synthesis its classification implies.

We note that the corpus of 30 papers passing all criteria across the years 1993–2024 is itself evidence of an emerging rather than an established intersection, which is the condition the Reviews and syntheses type is intended for (Sect. 2.2, p. 6, lines 153–157).

The subject of the minimum detection threshold or limit is widely mentioned, but I cannot see key references on it like the ones I add below. In these studies, they talk about how the detection limit of an area flux mapper depends on parameters such as pixel size, wind speed (related to the escape time of the gas inside a pixel) and revisit time.

Jacob, D.J. et al., 2022: Quantifying methane emissions from the global scale down to point sources using satellite observations of atmospheric methane.

Dubey et al. 2023: Minimum detection limits of the TROPOMI satellite sensor across North America and their implications for measuring oil and gas methane emissions.

We are grateful to the referee for these references, which have proved central to the revision. Both have been added and now underpin a new paragraph in Sect. 3.3.2 (p. 13, line 319-323). Jacob et al. (2022) supplies the scaling framework in which the minimum detectable emission rate depends on the product of pixel dimension and wind speed, the escape-time argument the referee identifies. Dubey et al. (2023) supplies the empirical quantification for TROPOMI: 500–8800 kg CH_4 h^{-1} per pixel for a single overpass, falling to 50–1200 kg CH_4 h^{-1} over a yearlong campaign, a reduction of roughly an order of magnitude driven by seasonal variation in wind speed and cloud cover. These figures now appear in Table 6 (p. 14) in place of the unsupported per-pixel threshold it previously carried, which also answers the referee's later question about that value.

Regardless of the articles above, the fact that the emission rates of small lakes and ponds are far below the current sensitivity of satellite approaches is already understandable in existing literature. My point is that there is no existing literature merging both topics because the sensitivity limitations from satellite instruments are rather well-stated.

We thank the referee for this framing, which states our contribution more sharply than we had stated it ourselves. Sect. 1.4 has been rewritten to open from this position (p. 5, lines 113–120): the sensitivity limits of satellite instruments are well established in the atmospheric remote sensing literature and the emission characteristics of small water bodies are well documented in limnology, and what has not been attempted is an assessment of the two together. The section now also asks for which size classes detection is feasible (p. 5, lines 117), which previews the analysis in Sect. 3.4.2.

You repeatedly discuss point-source imagers and their detection limit. It is of course important to define the limits of this approach. However, the problem you are researching is not a point-source but an area source. Thus, this technology is neither designed nor intended for your source. You point it out well in the article, but I believe it is too much of a central part of your discussion, when perhaps it should not.

We thank the referee for this critique, which identified a genuine structural weakness. The submitted manuscript never classified small water bodies as area sources: it classified sensors by architecture but not sources by geometry, which left the extended treatment of point-source imagers reading as an evaluation of a candidate technology rather than the demarcation of a boundary. Sect. 3.3.2 now states that small water bodies emit as area sources, that methane

leaves the entire water surface rather than a concentrated vent, and that even ebullition hotspots are distributed rather than co-located (p. 13, lines 312–315). It gives the physical consequence, that a diffuse source of a given total rate produces a smaller per-pixel enhancement than a point source of equal rate, and concludes that point-source imagers are therefore not the architecturally relevant comparison. The sentence beginning "This two category framework highlights why…" (preprint p. 12) has been removed as superseded. The area flux mapper treatment is correspondingly expanded with the Jacob and Dubey material above (p. 13, lines 319–323), and Sect. 4.1 recommends against point-source imagers for this source category (p. 18).

On a positive note, I enjoyed more the research and discussion on the optical proxies. It seems that, along with modelling and validation from in situ measurements, it would be the most likely option to succeed in bridging both disciplines. Thus, I perhaps would have liked to see a larger focus on this topic. The points you make in the discussion sections related to it are good. If you indeed believe optical proxies are the best option, the question that comes to my mind is: Given X spatial resolution or instrument characteristics, which percentage of the lakes and ponds of your dataset could be monitored?

We thank the referee for this question, which has produced what we now regard as the most useful result in the paper. The analysis has been performed on the full GSLPI inventory (n = 261,567) and is reported in Sect. 3.4.2 (p. 15–16, lines 359–366) and Table 7 (p. 16). Applying the criterion that a water body must span at least 3 × 3 pixels for retrieval free of adjacency effects:

Sensor	**Resolution**	**Minimum resolvable area**	**Water bodies resolvable**	**Share of inventory**
GOSAT/TANSO-FTS	~10 km	90,000 ha	0	0 %
TROPOMI (SWIR)	5.5 × 7 km	34,650 ha	0	0 %
PRISMA, EnMAP, Landsat-8	30 m	0.81 ha	31,471	12.0 %
Sentinel-2 (SWIR bands)	20 m	0.36 ha	63,080	24.1 %
Sentinel-2 (VNIR bands)	10 m	0.09 ha	184,106	70.4 %

The decisive result is one we had not anticipated: optical proxy retrieval does not use the shortwave-infrared bands. Chlorophyll-a and turbidity are retrieved from Sentinel-2's visible and near-infrared bands at 10 m, whereas the SWIR bands required for methane retrieval operate at 20 m (p. 15, lines 363–364). On the same platform, the proxy bands resolve almost three times as many water bodies as the methane bands. This quantifies the referee's intuition: the proxy pathway is not merely more promising in principle, it is the only pathway with substantial coverage of the inventory.

The percentages depend on the 3 × 3 criterion, which is stated as an assumption in the table note. The finding that methane-specific sensors resolve none of the inventory holds under any criterion, since TROPOMI requires 3,850 ha even at a single pixel.

Since you are reviewing these different methods, I would appreciate seeing more discussion of the specific potential of each of them, should the research be funded.

We thank the referee for this suggestion, which we have adopted in full. A new subsection, Sect. 4.1 Research priorities and value for limnological practice (p. 18-19, lines 431–465), treats each method class in terms of what funding would buy and what the binding constraint is. Optical proxy retrieval from Sentinel-2 is the most immediately practicable, limited by

calibration transferability rather than sensor performance, and the priority is a multi-site campaign pairing water-column CH_4 measurements with concurrent acquisitions across the trophic gradient (p. 18-19, lines 431–465). Emergent macrophyte extent is identified as a second and possibly more tractable proxy. Atmospheric inversion is the only route to a satellite-informed national estimate but cannot be resolved to individual water bodies, and the averaging that lowers its detection limit removes the resolution needed for episodic ebullition. Hyperspectral retrieval of submerged vegetation and suspended matter proxies substrate supply rather than methane and reaches only about 12 % of the inventory. Point-source imagers are not recommended.

I understand that this study is focused on literature reviewing, so I would not expect you to implement any method. But I feel that as it is, it does not add substantial value to the discussion of limitations and potential uses of satellite RS to the field of Limnology.

We thank the referee for this assessment, which we read as being about audience rather than about missing analysis, and accept on those terms. The submitted manuscript was written from the sensor side: it established what instruments can do and where they fall short, but did not state its findings in a form a limnologist could apply to a specific site. Sect. 4.1 now opens by restating the findings as size thresholds (p. 18-19, lines 431–465): below approximately 0.09 ha no satellite product is available at pixel level; between 0.09 and 0.36 ha chlorophyll-a, turbidity and emergent vegetation extent are retrievable from Sentinel-2's 10 m bands but no SWIR product is; above approximately 0.81 ha hyperspectral retrieval of submerged vegetation and suspended matter becomes possible; and at no size within the range is methane itself observable. Regional aggregate inversion is treated separately, since it constrains a budget total without attributing to individual water bodies. A reader can therefore determine from the surface area of a site alone which satellite products are worth acquiring.

Specific comments

Page 3, Line 70-75: Line 62 already states the origin of the data for Figure 1. I would skip detailing how the plot was produced and only add necessary information in the figure's caption.

We thank the referee for this suggestion and have adopted it. The description of the Python workflow has been removed from the main text (preprint p. 3, lines 70–75), and the figure caption retains the data source, the size class boundaries and the counts per class.

Page 5, Line 124-126: It's unfair to say there's a scarcity of papers dealing with the subject, when it's not even feasible to look at small water bodies to current sensitivity of satellite instruments.

We thank the referee for this correction, which is well taken. The passage now attributes the small corpus to the technological infeasibility of the problem rather than to limited research interest (p. 6, lines 153–157), and the sentence describing the field as having a thin literature base has been deleted (preprint p. 5, lines 124–126). Sect. 2.1 additionally states that the search concepts contained no plant-related terms and that wetland studies without a pond component were excluded (p. 6, lines 146–148), so the boundaries of the corpus are stated rather than inferred.

Page 9, line 201-202: Shouldn't the units of total emission rates and fluxes be the same in the description?

We thank the referee for pointing this out. The notation in that passage was inconsistent, giving kg CH_4 hr^{-1} in one instance and kg hr^{-1} in another, with a typographical error in the superscript.

Notation is now standardised throughout to kg CH_4 h^{-1} for emission rates and mmol CH_4 m^{-2} d^{-1} for areal fluxes, following SI convention, and the relationship between the two quantities is stated explicitly before the conversion in Eqs. (1)–(3) (p. 10, lines 239–244).

Page 10, line 215-220: I would recommend the last sentence right after mentioning "detection limits" in line 215.

We thank the referee for this suggestion, which we have adopted: the forward pointer now follows the detection-limit statement directly (p. 11, line 259), and a duplicate pointer has been removed (preprint p. 10, lines 215–220). In making this change we identified an arithmetic inconsistency the referee had not raised, and we are glad to have found it. The stated range of 10^3–10^4 did not span the values in Table 3, where Crampton Lake at 8.67×10^{-5} kg CH_4 h^{-1} falls 5.8×10^5 to 3.5×10^6 below the threshold band. The figure has been corrected to 10^3–10^6 (p. 11, line 259), and the Table 3 caption now states that the values are scaled to a 0.5 ha pond (p. 10).

Page 11, Table 4: The title is a bit misleading. The algorithms are not specifically used for aquatic environments, but for any other application.

We thank the referee for this correction. The caption now reads "Comparative analysis of satellite methane retrieval algorithms," and a table note states that none of the algorithms were developed specifically for aquatic environments, with a pointer to the section where their applicability to small water bodies is assessed (Table 4, p. 11).

Page 11, Table 4: The spatial resolution of the matched filter algorithm is sensor and platform dependent, not ~30m as stated.

We thank the referee for this correction, which identified a conceptual error rather than a numerical one. As the referee notes, the resolution is inherited from the host instrument and is not a property of the algorithm. The cell now reads "sensor- and platform-dependent," with footnote e giving the range: 3–8 m for airborne AVIRIS-NG, 20 m for Sentinel-2, 30 m for PRISMA, EnMAP and Landsat-8, and 60 m for EMIT (Table 4, p. 11). We also removed the word "sub-pixel" from that cell (preprint Table 4, p. 11), which conflated detection capability with resolution.

Page 11, Table 4: The output of the Remotec algorithm is total column mean-mixing ratio or column-averaged concentrations [ppb].

We thank the referee for this correction. The cell now reads "column-averaged dry-air mixing ratio, XCH_4 [ppb]," following the terminology of Lorente et al. (2021), and units have been added to the other cells in that row for consistency (Table 4, p. 11).

Page 11, Table 4: Daily global coverage is not a processing speed.

We thank the referee for spotting this. The row is now labelled "Computational demand" and reports relative computational cost across the four algorithms; daily global coverage has been moved to the detection scale and coverage row (Table 4, p. 11). Following the referee's point we also corrected a second instance of the same category error: the CO_2 proxy cell in that row previously read "near real-time," which is likewise a delivery-latency statement rather than a processing characteristic.

Page 11, Table 4: Specify that the quantity stated for point sources in the detection scale row is a sensitivity threshold.

We thank the referee for this clarification. The cell now reads "point sources; sensitivity threshold 50–300 kg CH_4 h^{-1}" (Table 4, p. 11).

Page 11, line 237: I disagree with the statement that radiative transfer models are sidestepped by the MF algorithm. RTM's can be used to generate a target spectrum under specific conditions, thus used along the MF algorithm (often).

We thank the referee for this correction, which is well founded. The original wording implied that matched-filter approaches operate independently of radiative transfer modelling, and the paragraph contradicted itself in doing so, since the following clause described the projection of radiance onto a target spectrum. The text now states that the matched filter avoids per-pixel radiative transfer inversion, while a radiative transfer model remains necessary to generate the target methane spectrum onto which each pixel's radiance is projected (p. 12, lines 284–287).

Page 12, table 5: Add EMIT and ENMAP to the point source imagers list, since you refer to them earlier.

We thank the referee for noticing this omission. EnMAP (Guanter et al., 2015) and EMIT (Thorpe et al., 2023) have been added to the point-source imager column (Table 5, p. 13). GHGSat, Carbon Mapper and MethaneSAT have also been added, which addresses Referee 2's related comment, and MERLIN is now cited to Ehret et al. (2017) rather than to sources that do not describe it.

Page 12, line 166: The matched filter algorithm, used by PRISMA and Sentinel-2 is not calibrated for any sort of target. The nature and sensitivity of the algorithm and the instrument can just detect concentrations above a certain threshold.

We thank the referee for this important distinction, which we accept. The text now states that the algorithm is not calibrated for a particular source type but flags any pixel whose methane enhancement exceeds the detection threshold set by instrument sensitivity and scene variance (p. 13, lines 287–288). The failure over small water bodies is attributed to its actual causes: dark water reflects little shortwave infrared radiation, degrading the signal-to-noise ratio, and diffuse emission produces a smaller per-pixel enhancement than a concentrated plume of equivalent total rate (p. 13, lines 288–290). The phrase "violate both design assumptions" has been removed (preprint p. 12, line 166), as it carried the same implication the referee objects to.

Page 13, line 267: The 3 citations in this line are more related to satellite remote sensing than eutrophic ponds. Their location is misleading since it seems they refer to the ponds.

We thank the referee for spotting this. The three citations supported the claim about PRISMA's detection limit but sat at the end of a sentence whose final clause concerned pond emission rates. The passage has been split into three statements so that Ayasse et al. (2019) and Guanter et al. (2021) attach to the detection limit, Jongaramrungruang et al. (2019) attaches to the conversion from column enhancement to emission rate, and the pond figures refer to Sect. 3.2.2 where they are derived (p. 15, lines 337–339). We also removed an order-of-magnitude comparison between a column enhancement in ppm m and a mass flux in kg h^{-1} (preprint p. 13, line 267), which are dimensionally distinct quantities.

Page 13, line 270-275: This has already been mentioned before. The pixel size mentioned here is different to the one stated in Table 4. Why? I believe 5.5x7 is the correct one.

We thank the referee for flagging this inconsistency, and for identifying the correct value. TROPOMI retrieves methane from the SWIR band, whose across-track ground pixel is 7 km, giving 5.5 × 7 km following the August 2019 along-track improvement; the 3.5 km across-track figure belongs to the UV–VIS–NIR bands from which NO_2 and ozone are retrieved. The value

in Table 4 was erroneous and has been corrected (p. 12), as have the further instances elsewhere, so the methane pixel is now stated consistently as 5.5 × 7 km (38.5 km²) throughout: p. 3, line 91; Table 4, p. 12; p. 14, line 331; Table 6, p. 15; and Table 7, p. 16. The derived figures of 38.5 km² and 0.013 % are likewise consistent. The repeated pixel calculation has been consolidated into Sect. 3.4 (p. 14). We also removed a citation to van Geffen et al. (2022) that had been attached to the methane pixel statement (preprint p. 13); that paper describes the NO_2 retrieval and does not support the figure, and it has been deleted from the reference list, where it was cited only once.

Page 13, line 276: Meter-scale resolution is not the same as 20–60-meter spatial resolution.

We thank the referee for this correction. The sensors named, Sentinel-2 (20 m), Landsat-8 (30 m) and PRISMA (30 m), operate at tens of metres. The text now states 20–30 m explicitly (p. 13, line 310; p. 14, line 334), and the further instances of "metre-scale" elsewhere have been corrected to give the actual figures.

Page 14, line 286-289: The emission values from small water bodies has been mentioned before. Phrasing it here does not add value to the narrative.

Page 14, line 289-298: Repetition of facts stated in the same section. They do not add value.

We thank the referee for these observations, both of which we accept. The emission-value list has been replaced by a single sentence referring to Sect. 3.2.2 (p. 15, line 348; the list removed from preprint p. 14, lines 286–289). The following paragraph, which restated the pixel dimensions, detection thresholds, emission rates and blind-test result already established earlier in the section, has been reduced to a lead-in sentence for the table plus the one point the table cannot convey: that the two constraints are independent, so neither sensor category compensates for the other's limitation (p. 15, lines 350–352; the restatement removed from preprint p. 14, lines 289–298). A duplicated statement of the Sherwin et al. (2024) result within the same section has also been removed (preprint p. 14, lines 289–298).

Page 14, line 301-302: I agree that the better instrumentation gets, the closer to tackling this problem we will be. However, the problem you are tackling would not be solved by a detection threshold of 10 Kg/hr or slightly below.

We thank the referee for this correction, which we accept without reservation. Against the values in Table 3, the minimum detection limits of 66–144 kg CH_4 h^{-1} predicted for Carbon Mapper (Duren et al., 2025) remain between roughly 4,000 and 1.7 million times above pond emission rates. The passage now states the sensitivity that would actually be required, noting that for the highest-emitting 0.5 ha case it lies roughly four orders of magnitude below that range, and further still for smaller or less productive ponds, and redirects the conclusion toward aggregate proxy-based approaches rather than implying that incremental instrumental improvement will close the gap (p. 14, lines 329–332). A second occurrence of the same claim in Sect. 3.5 has been corrected in the same way (p. 15, lines 354–356).

Page 14, Table 7: The detection threshold for area flux mappers is rather dynamic. What is the basis for the ~15 kg/pixel value?

We thank the referee for questioning this value; the query was well placed and the referee is correct on both counts. Area flux mappers have no fixed point-source detection threshold, since sensitivity depends on pixel size, wind speed, surface albedo, and whether detection is attempted in a single overpass or through multi-pass inversion. The ~15 kg/pixel value was not supportable: at TROPOMI's ~1 % XCH_4 precision, a detectable column enhancement over a 38.5 km² pixel corresponds to several thousand kg of CH_4 rather than 15 kg, and the quantity is in any case a mass rather than a rate. The table, now Table 6 (p. 16), reports the empirically

derived limits of Dubey et al. (2023), 500–8800 kg CH_4 h^{-1} per pixel for a single overpass and 50–1200 kg CH_4 h^{-1} for a yearlong campaign, which also answers the referee's earlier comment on missing detection-limit references. We additionally corrected the attribution of the 50–300 kg CH_4 h^{-1} range in the adjacent cell, which derives from Cusworth et al. (2018) and Varon et al. (2021) rather than from Sherwin et al. (2024), who report 60–200 kg CH_4 h^{-1}.

We should also record a change the referee did not raise. The submitted version contained two tables with the same caption, "Comparative analysis of satellite methane retrieval algorithms": Table 4 and Table 6. The latter was an uncorrected earlier draft of the former and has been deleted in revision. The former Tables 7 and 8 are therefore now Tables 6 and 7, and all cross-references have been updated.

Page 15, line 331-333: I would phrase this statement differently. It's not so much the design of the sensor, but the sensitivity of the technology.

We thank the referee for this suggestion, which we have followed. The passage attributed the detection gap to design intent. It now states the two physical limits directly: the methane enhancement produced by a pond of a few hectares falls below the detection threshold of every operational sensor, and within kilometre-scale pixels it is diluted below the retrieval noise floor (p. 17, lines 399–401).

Page 16, line 368-373: I agree with the first statement, that optical proxy retrievals are a promising option. Thus, it feels off to bring back point source imagers to the discussion.

We thank the referee for identifying this inconsistency. Having established optical proxy retrieval and statistical upscaling as the tractable near-term pathway, the passage then made progress conditional on improved point-source sensor sensitivity. The requirement for sub-10 kg CH_4 h^{-1} detection thresholds has been removed here and in Sect. 3.4 (preprint p. 16, lines 368–373), the remaining dependency is stated as systematic field calibration of chlorophyll-a–methane relationships across the trophic gradient, and the passage has been reordered so that the hybrid approach is described before it is referred to (p. 15, lines 354–356).

Technical corrections

We are very grateful to the referee for compiling this list, which was considerably more useful and all have been applied.