We thank the reviewer for their constructive comments and insightful suggestions, based on which we have thoroughly revised the manuscript. The main revisions are summarized below, with detailed, point-by-point responses provided in the following sections. Our responses are shown in black, while the original comments are reproduced in blue italics.

1. **Expanded false positive and false negative analysis.** We added a clear definition of a plume in the Methods and conducted a more comprehensive error analysis in the Results. We identified background interference and diffuse enhancements as the primary sources of false positives, and small plumes and low–flux emissions coexisting with strong sources as the main causes of false negatives, with additional challenges coming from merged plumes. These findings motivate an expanded discussion of future work in Discussion, including targeted wavelet denoising in high-probability subregions, adaptive and region-specific parameter tuning, and the development of classifiers for false detections.
2. **Broader comparison with alternative methods.** We extended the evaluation to include additional denoising approaches (Gaussian filtering and median filtering) and alternative wavelet families (Daubechies, Symlets, and biorthogonal wavelets), both assessed based on detection performance. We also added a comparison with a machine learning approach in terms of computational efficiency.
3. **Quantitative assessment using a probability of detection framework**. We incorporated a probability of detection (P_d) analysis following Manninen et al. (2026), enabling a more rigorous comparison of detection limits across detection methods and observing systems. Results indicate that the wavelet method generally outperforms DI thresholding by detecting more plumes, including those with lower emission rates. For MethaneAIR, the wavelet method achieves $P_d \approx$ 0.93 and 0.42 at wind speeds of 1 m/s and 10 m/s, respectively, for emissions of 1000 kg/h; for MethaneSAT, $P_d \approx 0.93$ and 0.43 under the same wind conditions for emissions of 2000 kg/h.

Major comments:

The manuscript presents a new method for masking plumes. The focus of the method, the wavelet transform is not strictly new, as it has been used before, both within and outside the field. Other spatial frequency-based techniques such as Difference of Gaussians, median filters, and other denoising techniques are also widely applied in the field. This undermining the claimed novelty of the method.

Response:

We thank the reviewer for raising this important point. We agree that wavelet transforms are not new in image processing, and the current manuscript lacks a comparison with other commonly used denoising techniques. We have revised the manuscript to clarify that the novelty of our work lies not in the wavelet transform itself, but in its application to plume detection tailored to high-resolution methane remote sensing data.

To address the reviewer's request, we conducted additional analyses comparing the wavelet denoising step with two alternatives: Gaussian filtering and median filtering. We directly replaced the wavelet denoising step with these methods while keeping all subsequent filtering steps identical. The comparison was performed across three MethaneAIR scenes and three MethaneSAT scenes spanning a range of background and wind conditions, consistent with those used in the sensitivity analysis.

The results (SI Table S8 and Table S19) show that the wavelet-based approach consistently yields a higher number of true detections compared to both Gaussian and median filtering. In addition, the alternative methods produce either comparable or higher numbers of false detections in most of the scenes. This indicates that while conventional denoising techniques can suppress noise, they are less effective at preserving plume structures across spatial scales, leading to reduced detection sensitivity.

We also evaluated the sensitivity of our results to the choice of wavelet by testing several commonly used families, including Daubechies (db4), Symlets (sym4), and biorthogonal wavelets (bior4.4). These wavelets are also widely used in image processing and represent different properties, such as compact support (db4), increased symmetry (sym4), and linear phase reconstruction (bior4.4). From the results we did not observe a significant difference between the Haar wavelet and the others in terms of the number of true and false detections (SI Table S9 and Table S20). In most cases, the other wavelets yield equal or fewer true detections and equal or more false detections. In addition, the plume masks generated using the Haar wavelet are generally similar to, and in some cases more physically reasonable than, those produced by the other wavelets (Figure S8). These results support our choice of the Haar wavelet as a simple and effective option for this application. We note that continuous wavelet transforms are not considered here as they do not naturally fit into this reconstruction-based denoising framework.

We have added subsections, Section 4.1.1 “Comparison with conventional denoising methods” and Section 4.1.2 “Sensitivity to wavelet choice”, into the Discussion to summarize these findings and clarify the advantage of the wavelet-based approach. We hope this additional analysis addresses the reviewer’s concern and strengthens the justification for applying wavelet transforms in plume detection.

While the authors show that their method is better than the method used in Warren et al., they still show that they miss about 10% of the plumes detected with the other method. Any discussion on why these plumes are missed is missing. Furthermore, it raises the question of the completeness of the plume detections from method presented here. The paper does not touch on this at all. I feel that a discussion on the number (and type) of missed plumes is required to properly assess the detection and masking method.

Response:

We thank the reviewer for raising this important point regarding missed detections and detection completeness.

First, we would like to clarify that the initially reported 86 false negatives in MethaneAIR data included cases due to differences between versions of the Level 3 (L3) methane concentration maps. The wavelet method was applied to an updated L3 dataset with a smaller spatial domain and additional screening, while the DI thresholding method used an earlier version with broader coverage. As a result, some “missed” plumes were not present in the wavelet input. After removing these cases, the number of false negatives decreases from 86 to 48. We have revised the manuscript to ensure a consistent and fair comparison.

We have further added a detailed false negative analysis in Results, which identifies the main factors contributing to missed detections. We find that in MethaneAIR results, 58% of missed plumes consist of only a small number of enhanced pixels, which are suppressed during the wavelet denoising step. 24% occur when weak plumes appear with strong plumes in the same scene, where the concentration filtering step raises the pixel value threshold and removes the weaker signal. Other less frequent cases include merged plumes being grouped into a single mask and plume fragmentation caused by upstream data screening. We have also plotted examples from each category in SI Figure S4 and Figure S5. A similar analysis for MethaneSAT shows only 2 false negatives across 23 scenes. Both cases involve weak plumes in the same scene with stronger plumes, consistent with the behavior observed in MethaneAIR (see SI Figure S7).

These observations suggest that the concentration filtering step needs to be improved by introducing adaptive or region-specific thresholds to better retain weak plumes in scenes with strong enhancements. Then regarding the small-size plumes, improvements to the wavelet denoising step could include less scale-dependent thresholding or targeted processing in high-probability regions to preserve compact plume signals while maintaining noise suppression (see more discussion in Section 4.3). Furthermore, we clarified in the manuscript that while the wavelet method misses some small plumes, the DI thresholding method miss a larger number of smaller plumes that the wavelet method finds.

Similar with what we've done for in the false negative analysis, we have also expanded the false positive analysis by adding explicit count proportions of each category and example figures (SI Figure S3 and Figure S6).

We agree that detection completeness warrants further discussion. In the revised manuscript, we first clarify that, currently in the absence of ground truth, completeness can only be assessed relative to a reference method. Then we discussed potential ways to improve completeness in Section 4.3 "Future work", including targeted processing of high-probability emission regions, adaptive and region-specific concentration filtering, and the development of artifact-specific classifiers. The first two directions aim to improve the sensitivity of the model, and the third direction (artifact classifiers) could allow for a balance between increasing true detections and avoiding false detections.

Furthermore, abstract and the conclusions focus on the wavelet transform as the main workhorse of the results. The Supplemental materials, however, show that the extra filtering steps are critical in lowering the number of false positives. How many false positives these steps exactly exclude is not shown. As such it is not clear if it is the masking on the wavelet transformed image that enables the results or if a better tuned version of the Warren et al method with the extra filtering steps would lead to a similar number of true plumes and false negatives.

Response:

We thank the reviewer for this important point regarding the respective roles of the wavelet transform and the subsequent filtering steps. To clarify this point, we have added a new table of three MethaneAIR scenes and three MethaneSAT scenes in the manuscript (Table 2) that report the number of true and false detections after each processing step.

As noted by the reviewer, the subsequent filtering steps, the concentration filtering in particular, play a significant role in reducing false positives. At the same time, these results also show that the wavelet denoising step determines how many true plumes are initially identified. In other words, while the filtering steps are critical for improving precision, the wavelet-based denoising and masking largely controls sensitivity. We have revised the manuscript to better reflect this complementary relationship.

Regarding the question of whether the DI thresholding method combined with the same filtering steps would achieve similar performance, there are several important differences. First, DI thresholding also incorporates a concentration-based filter that reduces false positives, but the number of remaining false positives is still higher than the wavelet method (SI Table S8 and Table S19). Second, the DI method operates on both XCH_4 maps and flux divergence maps, thus produces plume candidates with fundamentally different spatial characteristics (e.g., larger, more square-shaped regions). As a result, the shape filtering developed for wavelet-derived masks is not directly applicable to DI outputs. Finally, while the wind direction filter could in principle be applied to DI results, this step contributes the least reduction in false positives and can sometimes remove true detections. This implies that its overall impact would be limited.

We have added clarifications in Results to better separate the roles of wavelet denoising plus masking and subsequent filtering. We believe these additions provide a clearer understanding of how the overall performance is achieved.

Another problematic, major claim is on the computational efficiency. While this is a recurring claim in the paper, there is no result that compares the computational power required to process the observations, especially when including the parameter exploration required to optimize the algorithm.

Response:

We thank the reviewer for pointing out the lack of quantitative support for the computational efficiency claim. To address this, we have added a discussion section (Section 4.1.3) and a table in the SI (Table S21) that compare computational performance across the wavelet method, the DI thresholding method, and a representative ML-based approach (see more details about the ML method in SI Section 1.3). These together report total runtime and peak memory usage for processing a single scene, averaged over the same set of observations.

The results show that the wavelet method does not provide an advantage over the DI method in terms of runtime or memory usage. However, because it does not require training data or a training stage, the wavelet method is still substantially more efficient than the ML-based approach. This is consistent with our claim that the primary computational advantage of the wavelet method is relative to ML methods.

We acknowledge that the reported values include only the cost of applying the optimized methods, not the cost associated with parameter tuning. The computational cost of parameter tuning varies across methods. For ML approaches, models that leverage pre-trained weights and transfer learning can significantly reduce tuning cost, whereas training from scratch can be computationally expensive. In contrast, in the wavelet method, the computational cost of parameter tuning scales approximately linearly with the number of parameter combinations tested, and each evaluation is relatively inexpensive. In addition, many parameters are physically interpretable, which limit the extent of tuning required. While some tuning is required for a different observing system, we provide recommended parameter ranges and default settings in the code, allowing users to achieve reasonable performance without extensive exploration.

We have added this computation efficiency comparison to the Discussion to reflect that the computational advantage of the wavelet method is primarily in avoiding the training data and stage required by ML approaches.

For this manuscript to be reconsidered I would like to see the following three things:

1. *An evaluation of the merit of wavelet transforms over other image processing or denoising*
2. *A discussion on the choice of the Haar wavelet decompositions out of the large array of frequency filtering techniques out there.*
3. *An analysis of the completeness of the proposed method with a discussion on the type of plumes that are not captured by the current algorithm.*

Response:

We thank the reviewer for summarizing the key points. Points (1) and (2) are addressed in our first response above, where we provide additional comparisons with alternative denoising methods and evaluate the choice of wavelet families. Point (3) is addressed in our second response above, where we include a detailed false negative (and false positive) analysis and expand the discussion on detection completeness.

Minor comments:

Line 3: Plume masking does not seem critical to source localization; plume detection would be enough for that.

Response: We agree with the reviewer that source localization does not necessarily require full masking. We have revised this sentence as below.

"*Efficient and accurate detection and masking of emission plumes are essential for localizing and quantifying, point-source emissions via remote sensing.*"

Line 11-12: "Its high sensitivity to low-volume emissions also enables a lower detection limit and provides a more comprehensive emission rate distribution." This claim is only present in the abstract, furthermore the paper does not include any quantification of the detection limit or the completeness of the emission rate distributions.

Response: We thank the reviewer for pointing out the lack of quantitative support for this statement in the original manuscript. First, we clarify that qualitative evidence for improved sensitivity to lower-emission plumes was already presented in the Results section. In the MethaneAIR analysis, we showed that most plumes detected only by the wavelet method occur in lower flux ranges, with the dominant emission rate distribution shifting from 600–1000 kg/h in Warren et al. (2025) to 200–600 kg/h when including additional wavelet-detected plumes. A similar trend was also reported for the MethaneSAT results. These observations suggest improved sensitivity toward lower emission rates.

To address the reviewer's request for quantitative evaluation, we have now added a more rigorous analysis of detection limits and detection completeness based on a companion study (Manninen et al., 2026, https://doi.org/10.5194/egusphere-2026-115). This study develops a unified probability of detection (P_d) framework to compare plume detection performance across different observing systems (including MethaneAIR and MethaneSAT) and detection algorithms (including the wavelet method and the DI thresholding method). The framework introduces a nondimensional "observability" predictor that combines emission rate, wind speed, pixel size, and gas concentration noise, and maps it to detection probability using logistic regression (here observability =log $\left(\frac{q}{\sqrt{anu}}\right)$, q = emission rate per unit area, a = pixel area, n = noise amplitude, u = windspeed). The analysis is based on ~80,000 synthetic plumes generated using Weather Research and Forecasting Large Eddy Simulations (WRF-LES), spanning a wide range of atmospheric conditions and noise characteristics, as well as ~62,000 scenes derived from controlled-release experiments using image processing methods.

The results show that the wavelet method generally outperforms the DI thresholding method by detecting more plumes overall, particularly at lower emission rates, and by being more robust to real-world observations with missing data. Quantitatively, the wavelet method achieves higher P_d than DI above an observability of ~1.3, with the largest performance gains at intermediate observability (~2.5, a SI figure with P_d as a function of observability is provided). At very low observability, the wavelet method shows slightly lower P_d, likely due to suppression of weak signals during multi-scale denoising. This behavior is consistent with the patterns observed in our MethaneAIR and MethaneSAT case studies.

The companion study also provides quantitative detection limits under different conditions. For example, for MethaneAIR, the wavelet method achieves P_d of 0.93 and 0.42 at wind speeds of 1 m/s and 10 m/s, respectively, for emissions of 1000 kg/h. For MethaneSAT, comparable P_d values (0.93 and 0.43) are achieved for emissions of 2000 kg/h under the same wind conditions. A summary of detection probabilities across different emission rates and wind speeds is provided in Table 3.

We have added a subsection, Section 4.2 "Probability of detection" under the Discussion to better quantify detection limits and completeness. We believe these additions address the reviewer's concern and provide a more complete basis for the claim.

Line 12-14: Computational efficiency of the method is not really discussed in the paper and significant platform specific tuning has gone into the method presented as such the claims in these lines seem overstated.

Response: We thank the reviewer for this comment, and we kindly refer the reviewer to our previous response where we provide additional quantitative comparisons of computational performance across methods.

Introduction: The definition of what is and is not a good plume or plume mask varies by application. For example, in figure S4 the authors discard a plume mask that to many readers would be a fine mask to use. A discussion on the type of plume/plume mask the authors are looking for would help the reader understand what the authors are trying to achieve.

Response: We thank the reviewer for pointing out that the definition of a "good" plume or plume mask can vary by application. We agree that clarifying our objective is helpful to interpretate the results.

To address this, we have added a new subsection, Section 2.2.1 "Definition of a "plume"" in the Methods. We defined a plume as a connected region of enhanced methane concentration that either contains a clear hotspot that enables emission source identification, or exhibits a shape that is physically consistent with transported emissions and allows for source localization.

Under this definition, we exclude diffuse enhancements that represent real emissions but cannot be reliably linked to a source. These enhancements can come from: (1) dispersed small sources whose individual emission rates are below the detection limit; (2) intermittent emissions where the source stops emitting before overpass, resulting in enhancements observed significantly downwind of the source location; and (3) fragmented plumes where turbulence or eddies break the plume into multiple detached clumps, and those spreading downwind cannot be reliably traced back to a single source. This explains cases such as the figure mentioned by the reviewer, where a mask that may appear acceptable for other applications is intentionally excluded here because it does not support source localization.

In addition, we expanded the Results section to discuss diffuse enhancements as a major source of false positives and to show that the wavelet-based method is more effective than DI thresholding at distinguishing source-localizable plumes from diffuse enhancements. Together, these revisions clarify both our plume definition and how it influences evaluation of the methods.

Line 29: see line 3 above

Response: We have rephrased this sentence as:

"The second step often requires defining the plume boundary (mask), which is also a crucial input for visualizing plumes, and *can also support localizing sources (origins)*".

Line 30: When talking about the background interference here, it would be good to provide some examples.

Response: We have rephrased this sentence as:

"However, the retrieved plume concentrations are not always markedly higher than the background, due to either low plume concentration enhancements or background interference (*such as upwind emissions and surface heterogeneity*)."

Line 51-52: There is an abrupt change in the text flow between these two paragraphs. Guide the reader more.

Response: We have added a sentence here for a smoother change of topics:

"*We present here an alternative method that is both computationally efficient and less dependent on training data, based on wavelet denoising.* The 2D discrete wavelet transform is an image processing technique commonly used for image denoising and compression..."

line 70 (and 294): The authors claim the method is broadly applicable to multiple platforms and species, while they only discuss a single species with two (closely) related platforms. Substantiation of this claim is required.

Response: We thank the reviewer for pointing this out. We agree that our analysis is limited to methane observations from two closely related platforms, and we have revised the manuscript to highlight that our method is potentially generalizable, but it has not been broadly validated. Specifically, we state that our method is based on feature extraction and signal-vs-background separation, which do not rely on any physical or chemical properties of methane. Since it targets generic characteristics of point-source plume, such as the spatial coherence and localized enhancements, this method is designed to be readily adaptable to (1) different observing systems for methane point source detection, and (2) other trace gases that exhibit plume-like behavior (e.g., NO_2, SO_2), provided appropriate tuning of parameters. In the Discussion we explicitly note that such extensions would require validation and parameter tuning, thus are beyond the scope of the present study.

Section 4.3: "*Our method is based on feature extraction and signal-vs-background separation, which do not rely on any physical or chemical properties of methane. Given that it targets generic characteristics of point-source plume, this method is designed to be readily adaptable to different observing systems for any trace gases that exhibits plumes from discrete emitters. These extensions would require validation and parameter tuning, thus are beyond the scope of this study.*"

line 78: For readers a range of swath lengths or scene surface areas would help in understanding the coverage of the used observations

Response: We have revised this sentence to:

“With its high spectral resolution (0.24 nm) and spectral sampling (0.08 nm), large swath width (220-440 km) *and along-track length (~220 km)*, and fine spatial sampling (110 m across-track × 400 m along-track)...”

line 82: The word sounding seems out of place here, furthermore it creates ambiguity on the scales at which data is masked. Is this based on the pixel grid, the instrument spatial sampling or something else?

Response: We thank the reviewer for this comment. “Sounding” is used in MethaneSAT Level-2 data processing to distinguish Level-2 pixels with the radiance pixels. To avoid ambiguity, we replaced it with “Level-2 pixels”:

“*Level-2 pixels (110 m×400 m)* with cloud contamination and low-quality *data* were masked during upstream processing *to create the L3 products*.”

Line 115-120: Wavelet transform should be linear, so subtracting the high frequency image from the input image should be the same as setting the high frequency components to zero. Looking at the implementation on github, an observed difference in behavior might be purely due to clipping to positive values in intermediate steps. This should be explained properly in the main text.

Response: We thank the reviewer for this insightful comment. We agree that for an ideal linear and orthogonal wavelet transform, subtracting the high-frequency image from the input image is mathematically equivalent to setting the high-frequency coefficients to zero and reconstructing the image.

However, this equivalence does not strictly hold in our implementation due to several non-linear operations in our framework. First, the pre-processing step modifies the input image through thresholding and reassignment of high-value pixels to a uniform value, which alters the distribution of wavelet coefficients and breaks strict linearity. In addition, the wavelet decomposition is performed to a finite level (defined as half of its maximum possible value), and boundary effects introduce further deviations from ideal orthogonality. Furthermore, after the subtraction step, we further apply soft-thresholding wavelet denoising, which introduces an additional non-linear shrinkage of coefficients. This step selectively suppresses residual noise while preserving larger coefficients associated with plume structure. As a result, the overall workflow is not strictly equivalent to a linear low-pass filtering operation.

We have clarified these points in Section 2.2.2 to better explain the observed differences and the role of non-linear processing steps.

“*We note that for an ideal linear and orthogonal wavelet transform, the key subtraction step of our denoising method would be equivalent to setting the high-frequency coefficients to zero and reconstructing the image. However, this equivalence does not strictly hold in our implementation due to multiple non-linear steps in this framework, including the pre-processing thresholding, decomposition to a finite level, and the subsequent soft-thresholding of wavelet coefficients. These together result in a structure-preserving denoising rather than a true low-pass filtering.*”

Line 121: Setting L to half its maximum value feels arbitrary, has this been explored?

Response: We thank the reviewer for raising this important point. We agree that the choice of the decomposition level *L* should be justified in more details.

In response, we conducted a sensitivity analysis of *L* using three MethaneAIR scenes and three MethaneSAT scenes (the same scenes used for other parameter sensitivity analysis). We evaluated six candidate values corresponding to fractions of the maximum possible decomposition level: $L = L_{max} \times [\frac{1}{8}, \frac{3}{8}, \frac{1}{2}, \frac{5}{8}, \frac{3}{4}, 1]$. For each value, we quantified the number of true and false plume detections.

The results show a consistent trend in which both true and false detections generally decrease as L increases. Smaller values of L (e.g., $1/8L_{max}$) may retain more high-frequency components, resulting in a large number of false positives. In contrast, larger values (e.g., L_{max}) tend to oversmooth the data and suppress plume signals, leading to missed detections. Across all tested scenes, $L = 1/2\ L_{max}$ provides the best balance between detection sensitivity and false positive control. Based on these results, we selected $L = 1/2\ L_{max}$ as a robust and empirically justified choice. We have added this sensitivity analysis to the SI Table S2 and Table S12 and clarified the rationale for this selection in the main text.

"In this process, the decomposition level L is a key parameter. *Smaller values of L may retain more high-frequency components, resulting in a large number of false positives. In contrast, larger values tend to over-smooth the data and suppress plume signals, leading to missed detections (SI Table S2 and Table S12).* To balance these effects, we set L at half of its maximum possible value."

Line 124: I am missing the context for the soft thresholding wavelet denoising, please elaborate and add some references.

Response: We have added context and references to the soft thresholding wavelet denoising:

"Finally, we apply soft thresholding wavelet denoising to the processed image for further denoising. *This process begins with a wavelet decomposition, followed by shrinking the wavelet coefficients based on their magnitude, such that the coefficients lower than a defined threshold (usually zero) are reduced to the threshold value, while larger coefficients are preserved (Donoho and Johnstone, 1998).*"

Line 131: It is unclear if the scaling factor is based on the original image noise or the denoised image, please clarify.

Response: The scaling factor is based on the denoised image, and the resulting mask is then applied to the original image. We have revised this sentence to clarify:

"First, in the denoised image, pixels above a threshold value are preserved, where the threshold is defined as the local mean plus a scaling factor (value: 1.5 for MethaneAIR, 1.75 for MethaneSAT) times the local standard deviation (again, the sensitivity tests of these parameters are discussed in the SI Section 1.1). Note that the threshold is determined on the basis of the local background *in the denoised image to adapt to spatial heterogeneity caused by surface properties, atmospheric conditions, and retrieval noise.*"

Line 133: An imaged based local background is used here, while the wavelet transform also contains information on the background that could be used to correct for large scale background variations. Why was a local background measurement chosen?

Response: We thank the reviewer for this comment. We would like to clarify that while the wavelet transform does decompose the image into low-frequency and high-frequency components, the former is not used as a direct estimate of the background for plume detection. This is because the low-frequency components depend on the choice of decomposition level and do not explicitly capture local variance well. On the contrary, the local background (mean and standard deviation) provides thresholds to adapt to spatial heterogeneity caused by surface properties, atmospheric conditions, and retrieval noise, thus is more suitable for the step of plume mask generation. We therefore use local background statistics to define detection thresholds. We have clarified this distinction in the revised manuscript.

"Note that the threshold is determined on the basis of the local background *in the denoised image to adapt to spatial heterogeneity caused by surface properties, atmospheric conditions, and retrieval noise.*"

Line 142: A note on which data product the hotspot detection would be good for clarity.

Response: The data product used here is the original concentration map. We have revised this sentence to:

"The concentration filter aims to remove false masks by finding a high concentration "hotspot" *in the original concentration map*, a smaller XCH_4 clump with elevated pixel values within the mask."

Line 147: It is mentioned that the pixel threshold value for the plume hotspot is adjustable, but its value is not discussed anywhere in the paper of the supplementary materials

Response: We thank the reviewer for pointing this out. We now added a section in the SI (SI Section 1.1) to explain how the pixel threshold value is adjusted for MethaneAIR and MethaneSAT observations.

Line 149: This looks to be a version of Hysteresis Thresholding. Could the authors comment of if it is, and if so cite the relevant work?

Response: We thank the reviewer for this observation. The described concentration filter using hotspot ratio indeed shares conceptual similarity with Hysteresis Thresholding, as both approaches apply a form of two-stage decision-making to retain strong signals while conditionally evaluating intermediate cases. We have revised the manuscript and cited relevant work.

"Masks with a hotspot ratio in between are further evaluated by the latter filters *(this is conceptually similar with Hysteresis Thresholding in Canny (1986))*."

Line 187: This presumably already includes wind information? Please clarify in the text.

Response: Yes, the angle needed to draw the growing boxes is collected either from meteorological data or calculated as the angle of the plume major axis. We have revised our manuscript to clarify this.

"Wind speed *and direction* from the meteorological product *at 80m height (HRRR or GFS, selected to represent the surface layer windspeed)* was used for this calculation, and for sufficiently elongated plumes (eccentricity > 0.87), the wind direction was rotated to match the angle of the major axis of the observed plume."

Line 193 "all plume detections presented in the Results section" Unclear phrasing: Is this limited to the plumes presented in Figure 3 and 6, or does it also include all plumes underlying the statistics in Fig 4 and 5?

Response: By "all plume detections" we mean the complete set of plume detections from both MethaneAIR and MethaneSAT, including both those explicitly plotted and those underlying the statistical results.

Line 216: The supplied version of the SI does not contain a full list of the 262 additional plumes.

Response: We thank the reviewer for noting this. Given its large size, the full list of the additional plumes was provided as a separate CSV file during submission, rather than being embedded within the SI PDF file. We will ensure it is properly included and linked as part of the Supplementary Information in the final submission.

Figure 2: The note in the figure implies that a number of flights have not been processed. From the figure this would be about 14% of the total DI thresholding plumes. It would be good to either include these flights in the analysis or include a discussion on why these flights are excluded.

Response: We thank the reviewer for this suggestion. We have now included these flights in the analysis, removed such statement from the manuscript, and updated Figure 2.

Figures 2, 4 and 5: The stacked bars feel like the wrong choice for the visualization here. The 200-400 kg/h bar in figure 2 can easily be interpreted as ~18 DI only plumes, ~25 DI + WL plumes and ~88 wavelet only plumes, while I think that is not the message of these figures. The use of irregular spacing of the x- and y-axis ticks and frequency as y axis label "frequency" also feel off.

Response: We thank the reviewer for this comment. Our primary motivation for using stacked bars is to show the total number of detections within each emission rate bin, thereby providing a view of the emission rate distribution. For example, the observation "dominant emission rate distribution shifting from 600–1000 kg/h in Warren et al. (2025) to 200–600 kg/h when including additional wavelet-detected plumes" is clearer in stacked bars. To avoid misinterpretation, we have added explicit annotations to one of the stacked bars, indicating the number of plumes detected by DI only, by both methods, and by the wavelet method only. In addition, we have revised the figures by standardizing the spacing of the x- and y-axis ticks and replacing the y-axis label "frequency" with "counts."

Line 225: Statement on future work is very vague and generic. It would be good to more concretely identify specific challenges and solutions.

Response: We thank the reviewer for this important suggestion. We have added Section 4.3 "Future work" under the Discussion to discuss specific challenges and potential solutions.

Specifically, we now clarify three specific challenges. First, the current wavelet framework is applied uniformly across entire scenes without using prior information on likely emission sources, which can limit sensitivity to weak plumes in high-probability regions. To address this, we propose targeted wavelet denoising to subregions using external source information, such as IMEO methane point source database and OGIM infrastructure data. This would allow more permissive thresholds in relevant subregions while maintaining control over false positives.

Second, we identify limitations in the concentration filtering step, which can suppress weak plumes in the presence of strong emissions and does not fully account for spatial variability in background. To address this, we propose adaptive and region-specific thresholding, potentially guided by a localized metric similar to the "observability" defined by Manninen et al. (2026), to improve robustness across heterogeneous scenes.

Third, we note that certain false detections arising from distinct artifact types, such as albedo-driven signals and diffuse enhancements, are not explicitly separated in the current framework. We therefore propose developing false detection classifiers that incorporate physically informed features such as plume morphology, surface reflectance patterns, and wind alignment.

Together, these revisions provide a more concrete set of challenges and corresponding directions for improving detection performance.

Line 232: Segmented plumes is a problem for most instruments, especially high-resolution instruments. The comment about this being a methaneAIR/methaneSAT problem feels out of place.

Response: We thank the reviewer for this clarification. We agree that segmented plumes are a challenge for many instruments, and we did not intend to suggest that this issue is unique to MethaneAIR or MethaneSAT. Our point is that diffuse enhancements (of which segmented plumes are one example) are especially challenging for MethaneAIR and MethaneSAT due to their high sensitivity and large spatial coverage, which results in frequent observations of long and gradually decaying downwind plume tails. We have revised this paragraph and the MethaneSAT results to clarify this point.

Section 3.1: "*Diffuse enhancements* are particularly challenging for MethaneSAT and MethaneAIR as a result of their high sensitivity and large spatial coverage, with clear signals observed well downwind that would not be detected by many other platforms."

Section 3.2: "We generally found more false detections for MethaneSAT than MethaneAIR using both the wavelet and DI thresholding methods, primarily because the high sensitivity of MethaneSAT led to more frequently observed diffuse enhancements, *which account for 65% of the total false positives. This is largely due to its ability to capture long plume tails (on the order of tens of kilometers), which, under complex wind conditions, can fragment into spatially dispersed enhancements that are difficult to trace back to a single source. In addition, the broader spatial coverage and higher signal-to-noise sensitivity of MethaneSAT make it more likely to detect intermittent emissions, further increasing the occurrence of diffuse enhancements.*"

Figure 3. The plume in panel b) seems like a well-defined plume that should not be hard to catch with a thresholding method. Why was this plume not captured originally?

Response: We thank the reviewer for this observation. While the plume in Fig 3 (b) appears visually well-defined, its detection using the DI thresholding method depends not only on elevated concentrations but also on the flux divergence signals.

The DI thresholding approach computes flux divergence over $600 \times 600\ m^2$ tiles across the scene, and candidate plumes are identified only where both elevated flux divergence and elevated XCH_4 values co-occur, and where the resulting clumps pass additional thresholding criteria such as the minimum clump size and proximity between two types of clumps (see Section 3.1). In this case, although the plume exhibits a clear concentration enhancement, the corresponding flux divergence signal is either too weak or not sufficiently co-located with high XCH_4 pixels to satisfy the joint thresholding criteria. As a result, the plume is not detected by the DI thresholding method.

We have clarified this point in the caption to better explain why this plume was not captured by the DI thresholding method.

Figure 3 caption: "MethaneAIR plume examples (background © Google Maps 2023). Compared to the DI thresholding method, the wavelet method efficiently captured more low-flux plumes like (b). *Plume (b) was not detected by the DI thresholding method as it did not satisfy the joint thresholding criteria requiring both elevated concentration and flux divergence.*"

Line 251: The concept of 'collection' is not introduced.

Response: Here "collection" refers to an individual MethaneSAT scene (a single observation swath), thus we changed it to "scene" to avoid confusion.

Line 265: In section 2.2.3 it is mentioned that plumes without clear heads are removed as possible plumes, but in these lines it it implied that they are still included as false positives.

Response: We thank the reviewer for bringing this point. Section 2.2.3 introduces the concentration filtering step, which is not perfect in practice. Some diffuse enhancements or weakly structured plumes may still pass the criteria and remain as detections.

This behavior is consistent with what we previously noted for MethaneAIR results, where "although many diffuse emissions were successfully filtered out by concentration filtering and mask size thresholding, some were still identified as plumes," and more generally that "although the wavelet method substantially reduces noise, it does not eliminate it." The same limitation applies here in the MethaneSAT results. To avoid confusion, we now explicitly clarify that our concentration filtering step needs to be improved, and other potential solutions include exploring adaptive and region-specific threshold tuning, and designing specific diffuse enhancements classifiers that distinguish them from true plumes and background (see Section 4.3).

Section 4.3: "*The current concentration filtering step does not fully address diffuse enhancements and can be influenced by the presence of high-emission plumes. Future improvements will explore adaptive and region-specific threshold tuning. For example, a localized metric similar to the "observability" defined by Manninen et al. (2026) could be used to determine pixel-level thresholds within each subregion. This metric would account for spatial variability in background heterogeneity, noise characteristics, and wind conditions, enabling more consistent performance across diverse scenes.*

Additionally, artifact identifiers false detection classifiers targeting specific artifact types can be developed and integrated into the plume detection models. For example, albedo artifacts and diffuse enhancements may be classified based on the spatial patterns of surface reflectance and methane concentration, as well as plume morphology and wind alignment. Incorporating these features into a post-processing classification step could reduce false detections while allowing more permissive plume-identification thresholds, thus improve overall detection sensitivity without sacrificing reliability."

Line 276-279: Similar to the statement in line 225 this is very vague and requires some substantiation.

Response: We thank the reviewer for this important suggestion. We have expanded these statements into a full "Future work" subsection under the Discussion to discuss specific challenges and potential solutions (Section 4.3).

Line 292-293: (see major comments) The computational cost of the wavelet method, especially optimizing it for a new platform or species is not discussed, to it is impossible to judge if it is more efficient than machine learning.

Response: Again, we thank the reviewer for pointing this out, and we kindly refer the reviewer to our previous response to the major comment.

Line 294: Same comment as for line 70.

Response: We kindly refer the reviewer to our previous response to the comment on line 70.

Line 297: It is very much appreciated that the authors share their code, however, for the audience it would be useful if it was expanded to include an example application of the workflow images of figure one or a feel reproduction package of the data used in the paper.

Response: We thank the reviewer for this helpful suggestion. We have now added a “test” folder to the repository that includes example input data (corresponding to Fig.1(a)) along with a script to run the workflow. This allows users to reproduce the output shown in Fig.1 and better understand how to apply the method in practice.