We are grateful to the referee for the very detailed feedback and interest in our work. Our answers to the comments and questions are provide below in red.

Comprehension

The study examines the feasibility of estimating methane emissions from the Nord Stream 2 (NS2) leak near Bornholm Island in September 2022 using Landsat 8 (L8) and Sentinel-2B (S2B) imager data in two bands of the short wave infrared spectral range. The authors utilize sea foam observations and employed the Multi-Band Single-Pass (MBSP) for the estimation of methane enhancements. For spectral calibration sea foam observations from ship trails are used. For the quantification of leak rates, they use the Integrated Mass Enhancement (IME) method, calibrated for their problem. It is found that even with these adjustments of the MBSP and IME, no firm conclusion can be drawn from individual L8 and S2B detections of the methane leak resulting in large uncertainties in the averaged leak rate estimate.

General comments:

In Section 2.3, please include an introductory sentence outlining the methods that could potentially be used for source rate estimation regarding the NS2 problem. After that, explain why IME was selected as the preferred method for quantification.

We added this introductory sentence and justified our choice of the IME.

New text: line 160 – 164	2.3.1 General description If a plume is observed in an image resulting from MBSP, the associated emission rate can be quantified using **different approaches such as the Gaussian plume inversion (GP), source pixel (SP), Cross-Section flux (CSF) and Integrated Mass En- hancement (IME) methods (Varon et al., 2018). Because GP and SP are not suited for the quantification of plumes detected using high-resolution satellite observations, and the CSF relies on several transects drawn on an extended downwind plume, we use the IME method.**

Please annotate the uncertainties discussed in Section 3 (as well as in the caption of Fig. 5) with the corresponding numbers from Section 2.4.

We understand this comment as asking to relate uncertainties obtained for methane leak rates to the width of the interval explored by the ensemble for each of the six input parameters.

Following Referee 2's comments, we modified this ensemble quantification to a full Monte Carlo ensemble quantification, thus considering the actual distributions of the input parameters that we explore to assess methane leak rate uncertainties. In addition to this new way of generating the ensemble, we also compute the first-order sensitivity indices for all six input parameters. They describe the contribution of each input parameter variance to the methane leak rate variance. The obtained first-order sensitivity indices answer this question of relating input parameter uncertainty to methane leak rate uncertainty.

The method that we employ to compute the indices is now detailed in an expanded and revised Sect 2.4, and the indices and conclusions that follow are given at the beginning of Sect. 3 in the revised manuscript. The Supplements now also include intermediate result plots illustrating the calculation of these indices.

Consider adding a table that displays the respective 'c' values for the MBSP calibrations. Alternatively, refer to the comments on figures in the specific comments section.

We understand that this general comment is related to the specific comment on Figures 3 & 4: "Please ensure, and specify in the caption, that the mean calibrations in Fig. 4 are based on all the ships listed in Tables 2 and 3, respectively", and that the table discussed here would include all 27 and 38 individual satellite-wise c values for L8 and S-2B, respectively.

We have expanded the Tables in the Supplements to include a column that provides the c calibration coefficient for each ship observation (which were already printed in the scatter plots included in the Supplements), and we have modified the text in Sect 2.2.2 and in captions of Fig 3 and 4 to refer the reader to the Supplements.

New text: line 135	[...] to determine c_i, the coefficient describing the spectral dependence of sea foam albedo for the i-th image (see individual **c_i values and fits obtained for each ship trail observation** in the Supplements).
New text: Fig. 3	Empirically determined sea foam albedo spectral dependence between s_1 and s_2 for Landsat 8 (left) and Sentinel-2B (right). Sea foam pixels for all ship images are depicted (**dots with different colors indicating different ships**, the legend only includes elements for the first and last images), along with their respective calibration slopes (**thin lines, each is detailed in the Supplements,** the legend only includes elements for the first and last images**.
New text: Fig. 4	Comparisons of s_1 and s_2 TOA reflectance (bottom) depicting different pixel types and showing the empirically determined spectral dependence of sea foam albedo (thick blue line, **the individual ship trail observations underlying this result are shown in Fig. 3 and in supplementary Tables S2 and S3 for L8 and S-2B, respectively)**, and the s_1/s_2 ratio observed over the NS2 sea foam patch (red line).

Specific comments:

Sec. 2.2:

It's imperative to immediately clarify that the standard approach for the MBSP isn't suitable for the NS2 problem.
We now included a short introduction to Sect 2.2 that already announces that the usual MBSP calibration will prove to be unsuitable for this specific NS2 case study. Symmetrically, we have also written a short introduction to Sect. 2.3.

New text: line 92 – 94	2.2 Methane enhancement retrieval: the Multi-Band Single-Pass (MBSP) method

We use the Multi-Band Single-Pass (MBSP) method to retrieve local methane column enhancements from Earth imager observations. We first describe MBSP and its standard calibration approach, and then show how this specific NS2 case study calls for a custom calibration. |

	2.2.1 General description
New text: line 156 – 158	**2.3 Emission rate quantification: the Integrated Mass Enhancement (IME) method**
	We use the Integrated Mass Enhancement (IME) method to quantify the methane emission rate from local methane column enhancement retrievals that show an emission plume. Here, we first explain why we choose the IME method and how it works, then we explain why this specific NS2 case study also calls for a custom calibration for the IME method.
	2.3.1 General description

In Fig. 1, it should be immediately evident that the variabilities within the L8 and S2B scenes, combined with the CH4-impacted sea foam pixels, are not suitable for MTSB calibration for CH4 detection.
The new short introduction to Sect. 2.2 now provides this information (see above). Besides, we have added this information in the caption of Fig. 1 as well, as recommended later in this review.

Moreover, given that we anticipate little to no methane signal from dark, still sea pixels, as suggested by Fig. 1, what is the rationale behind including those pixels in the fit for the linear calibration coefficient 'c'?
At the stage of Figure 1 in the paper, the rationale is to illustrate the "standard" naïve MBSP calibration strategy by actually applying it as it was presented in Varon et al. (2021) on the data at hand for L8 and S-2B, meaning on the full image (as it was presented), and not just the sea foam pixels. We have changed the text to better reflect this rationale.

New text: line 123 – 125	This issue similarly applies to the L8 NS2 observation, that also features an additional complication: very bright clouds are present in the image, which in this case drive the standard MBSP calibration (c = 1.13). **Thus, the standard MBSP calibration lines included in Fig. 1 illustrate why t**he NS2 observation case, that relies on a small sea foam patch, calls for an external calibration of the spectral dependence of sea foam albedo.

The term "standard calibration" might be misconstrued. Perhaps consider an alternative term, such as "naïve calibration"?
We agree that the standard calibration employs a naïve approach, and have included this adjective in a few places in the revised manuscript to describe it in Sect 2.2.1. In addition, we also specified that it will be "hereafter referred to as 'standard MBSP calibration'".

New text: line 102 – 105	**This calibration strategy was proposed with the MBSP method by Varon et al. (2021),** and implicitly assumes that image-wide pixels are representative of the surface characteristics expected below the (potential) methane plume. **Hereafter, we will refer to this "naïve" calibration strategy as the "standard MBSP calibration".** The rationale of MBSP is that deviations in the methane-sensitive s2 band [...]

In my opinion it's crucial to be upfront about the definition and shortcomings of the standard MBSP calibration with respect to this study.

The short introduction that we added to Sect. 2.2 addresses this question of being upfront about the shortcomings that are going to be described in the subsection (see above).

Sec. 2.2.1:
In my assessment, upon reading the section, it immediately becomes evident that the assumption of image-wide pixel calibration, representative for the surface characteristics beneath the plume, is untenable for the context of this study. It should be highlighted right away.
The short introduction that we added to Sect. 2.2 addresses this question of highlighting right away that the standard MBSP calibration will prove unsuitable for this specific NS2 case, with explanations given later in Sect. 2.2.2 (see above).

Fig. 1:

Enhance the caption with more detailed information.
We revised the caption to provide more detailed information.

New text: Fig. 1	Landsat 8 (left, **September 29th 2022**) and Sentinel-2B (right, **September 30th 2022**) images of the Nord Stream 2 leak for s1 (top), and s1 and s2 TOA reflectance comparisons depicting different pixel natures and showing the standard MBSP c calibration line (bottom). **The pixel natures of dark still sea (black), clouds (grey) and NS foam patch (red, all influenced by the methane leak) are separated using empirically determined thresholds given in the Supplements. The standard MBSP calibration (dashed line) is provided here to illustrate why it proves to be unsuitable for this specific NS2 case, as detailed in Sect. 2.2.2.**

It needs clarification that, without adjustments tailored to the NS2-specific challenge (CH4-contaminated sea foam over dark water pixels), the default MBSP calibration falls short of being appropriate.
We have elaborated the Fig. 1 caption following this comment. It now explains that this standard calibration will prove unsuitable for this case and refers to the discussion of this point in Sect. 2.2.2 (see above).

It's worth noting that no background (CH4-free) sea foam pixels are present in the target scene, as depicted in Fig. 1.
This point is indeed discussed in Sect. 2.2.2 when describing why the standard calibration is unsuitable for this case study. We added an element to the caption of Fig. 1 to reflect this aspect: "NS foam patch (red, all influenced by the methane leak)", see above.

The inclusion of the bottom row of Fig. 1 might be redundant since Fig. 3 already encapsulates that information.
Information is slightly redundant indeed, but Figures 1, 3 and 4 have been designed to follow a progression to better underline the different steps of the work we perform.
  - Figure 1 provides a first candid look at the data, thus includes the naïve standard calibration which is now commented upon in the caption and helps to clarify why this standard approach is not suitable for this case.

Furthermore, the lower panel of Fig. 1 primarily demonstrates an incorrect calibration method for the given context. If it's retained, the caption must be considerably elaborated.

Fig. 1 & 4:

Merging Fig. 1 and Fig. 4 (bottom rows, respectively) into a singular, per-satellite, introductory figure might be a viable approach?

Fig. 3 & 4:

Please ensure, and specify in the caption, that the mean calibrations in Fig. 4 are based on all the ships listed in Tables S2 and S3, respectively.

New text: Fig. 4	Comparisons of s1 and s2 TOA reflectance (bottom) depicting different pixel types and showing the empirically determined spectral dependence of sea foam albedo (thick blue line, **the individual ship trail observations underlying this result are shown in Fig. 3 and in supplementary Tables S2 and S3 for L8 and S-2B, respectively)**, and the s1/s2 ratio observed over the NS2 sea foam patch (red line).

Furthermore, clarify the rationale behind showcasing ships 1 and 27. Are they particularly unique, or are they simply randomly selected references?

New text: Fig. 3	Empirically determined sea foam albedo spectral dependence between s1 and s2 for Landsat 8 (left) and Sentinel-2B (right). Sea foam pixels for all ship images are depicted (**dots with different colors indicating different ships, the legend only includes elements for the first and last images)**, along with their respective calibration slopes (**thin lines, each is detailed in the Supplements, the legend only includes elements for the first and last images).**

Fig. 5:

Consider to add P(Q) in the caption.
We added the P(Q ≤ 0) notation explanation in the Fig. 5 caption.

New text: Fig. 5	Distributions of methane emission rate values for the Landsat 8 (left) and Sentinel-2B (right) ensembles. **Monte Carlo** ensemble means and standard deviations are shown inset, along with the fraction of null or negative emission rates, **denoted as P (Q ≤ 0).** The color scale shows the contributions of different sea foam albedo spectral dependence calibration values to the overall distribution of leak rates within the ensemble.

Fig. 2:

Were the dark sea and ship pixels also excluded from the analysis based on the tables provided in the appendix? Please incorporate this detail into the caption.
Yes, they were, as detailed in the text (line 123 in the original manuscript). We have added this explanation in the Fig. 2 caption as well in the revised manuscript.

New text: Fig. 2	Example of sea foam observation in the Sentinel-2B image of a ship trail acquired on October 12th, 2022. Dark still sea and ship pixels have been removed and are shown in grey and white, respectively. **They are also excluded from the sea foam albedo spectral dependence results presented later in Sect. 2.2.2 and in the Supplements.**

Fig. 4:

The elevated slope of the NS2 leak patch in Fig. 4, in comparison to the mean calibration from ship foam, seems to be primarily influenced by the notably bright s1 values. This observation becomes more apparent with the distribution of red dots around the fitted red line for lower values; they appear evenly dispersed, and in some instances, seem closer to the blue line. If this observation is accurate, it would be beneficial to note in the caption. This trend could suggest that source attribution might only be feasible for a select number of extremely bright spots (possibly bubbles?), where the Signal-to-Noise Ratio (SNR) is sufficiently high to discern the CH4 enhancement.

Indeed, the difference between the ship trail-based calibration and the s1/s2 fitted line on NS2 sea foam pixels is less apparent for the lowest sea foam albedo pixels in the NS2 images. When we get closer to the center of foam patch, it becomes brighter, thus potentially enabling to better distinguish the absorbing impact of methane on the s2 band, which leads to lower s2 values than the empirical calibration (blue) line, that thus results in pixels being above (and left) of the ship-based calibration (blue) line. So, we agree with this "higher SNR" interpretation of the "brighter foam patch center", and have extended the Fig. 4 caption to include this idea.

New text: Fig. 4	Comparisons of s1 and s2 TOA reflectance (bottom) depicting different pixel types and showing the empirically determined spectral dependence of sea foam albedo (thick blue line, **the individual ship trail observations underlying this result are shown in Fig. 3 and in supplementary Tables S2 and S3 for L8 and S-2B, respectively)**, and the s1/s2 ratio observed over the NS2 sea foam patch (red line). **The higher slopes shown by the s1/s2 ratios (red) compared to the empirical calibrations (blue) are driven by the brightest pixels at the center of the sea foam patch that offer a better signal-to-noise ratio to observe methane absorption than darker pixels.**

l 135-140 (Fig. 3 & 4):

Given the close relation between the bottom plots of Fig. 3 and Fig. 4, it might be prudent to present them within a single figure, divided into four subplots.

Please refer to the answer above regarding the progression between Figures 1, 3 and 4.

l 97:

Following the statement that the calibration strategy implicitly assumes that image-wide pixels are representative of surface characteristics, it's crucial to note that such an assumption is not valid for this particular problem.

The purpose of Sect 2.2.1 is to describe MBSP in the general case, as it was first presented in Varon et al. (2021). This comment has been addressed by including a short introduction to Sect. 2.2 which announces that the standard calibration of MBSP will prove to be unsuitable for this NS2 case (see above).

l 100:

Please provide some more details on the compilation process of the pre- computed look-up table? Additionally, it would be helpful if you could provide a reference to the radiative transfer (RT) code or the specific table employed.

We have added an extra sentence describing the input atmosphere and spectroscopic database (HITRAN 2020) that we employ to generate the look-up-tables.

New text: line 108 – 110	The translation of ΔR to methane enhancements is performed using pre-computed look-up tables, generated through radiative transfer simulations. **Here, they are based on the 2020 version of the HITRAN spectroscopic database (Gordon et al., 2022), rely on a 21-layer atmospheric model representative of mid-latitudes and include the impact of the solar zenith angle.**

l 137:

Perhaps the term "ship foam" should be placed somewhere to remind readers that the average empirical calibration was derived from ship foam observations. Consider to modify the statement to: "... the negative difference of the mean to the ship foam pixel calibration ...".

We adjusted the sentenced in the revised manuscript as suggested.

New text: line 148	This **ship-based $\bar{c}$ −s1/s2** negative difference overall translates to positive methane enhancement through MBSP.

l 157:

It might be beneficial to mention why Ueff also varies based on the type of observer, especially for Earth-like imagers.
This is already implicitly mentioned in line 156 of the original manuscript, when explaining that LES simulations have to be resampled according to instrument characteristics (spatial resolution, noise, etc). We have added an extra sentence to better reflect this comment.

New text: line 174 – 175	Plume transport includes complicated three-dimensional and turbulent effects that require computer-intensive simulations to be accounted for, if even possible given the randomness of turbulence. Through IME, the overall impacts of those effects are presumably captured into a single effective wind speed, denoted Ueff. Ueff is calibrated against the 10-m wind speed provided by meteorological models (U10m) over a set of Large Eddy Simulations (LES) made for known synthetic emission rates, and re-sampled according to a given instrument characteristics (spatial resolution, noise model, etc.). **Thus, Ueff can be calibrated for specific instruments and observing conditions.** Varon et al. (2021) provide an effective wind speed calibration model for Sentinel-2-like Earth imagers: Ueff = 0.33 × U10m + 0.45.

l 207-208:

You choose 10% because the fraction of negative emissions is roughly 10%?
10% was rather an arbitrary symbolic threshold with no justification. We do not employ it anymore in the revised manuscript (see revised manuscript Sect. 3).

l 203-204 & l 212:

Are you suggesting that the primary source of uncertainty stems from the uncertainties inherent in the imager's observations?
The primary source of uncertainty is the uncertainty on the spectral dependence of sea foam albedo. In the revised manuscript, it is now clearly shown thanks to the calculation of first order sensitivity indices. In line 212 in the original manuscript, we report "uncertainties" (methodological drawback may be a more appropriate expression) as written in Jia et al. (2022).

l 212-214:

A sentence for the conclusion?
A similar message in developed in a longer piece of text in the conclusion indeed. We think this sentence is relevant here as part of the discussion to explain that we explored the methodological drawbacks acknowledged by Jia et al (2022).

l 216-217:

It would be beneficial to elaborate further on the statement in parentheses, specifically explaining the reasoning behind the inability to assume independence.

We adjusted the revised manuscript to develop the reason why the quantifications may not be independent in a sentence before this one. The reason is that both satellite observations are processed with Look-up-tables that can for example be hampered by similar spectroscopy error originating from the HITRAN 2020 database itself, or by errors coming from the fact that IME estimates rely on the same set of LES simulations.

New text: line 258 – 259	[...] have ±1σ uncertainty intervals that include zero emissions, and show P (Q ≤ 0) = 0.20 and P (Q ≤ 0) = 0.21, respectively. **These separate L8 and S-2B estimates may not be independent. For example, similar look-up-tables or IME effective wind calibration errors or biases may hamper them.** However, if we opportunistically assume that they are, we can generate an ensemble [...]

l 218-220:

It would be beneficial to elaborate further on the statement in parentheses, specifically explaining the reasoning behind the inability to assume independence.

Please refer to the answer to the previous item.

l 218:

What does 1M stand for?

It stands for 1 million, we stopped using this notation in the revised manuscript.

Appendix, Table 1:

How is cloud classification defined for S2B? This is crucial, especially considering there will definitely be ship foam pixels where accurate calibration is important.

There is no "cloud classification" performed for S-2B as there are no clouds to remove from this image of the NS2 leak acquired by S-2B on Sept. 30th 2022. We adjusted the table to state that no cloud filtering is needed for S-2B.

New text: Table S1	**No cloud filtering required for this S-2B image**

All sea foam images in ship trails have been chosen so that cloudy pixels are not included in the s1 against s2 fits. This can be easily verified by examining Figures included in the Supplements: the vast majority of the pixels (dots) satisfactorily align with a s1/s2 = 1.8-2.0 slope, which is far from the cloud-related s1/s2 = 1.13 slope shown by the standard L8 calibration that was driven by cloudy pixels before the empirical ship-based calibration (see Fig. 1).

References

Varon, D. J., Jervis, D., McKeever, J., Spence, I., Gains, D., and Jacob, D. J.: High-frequency monitoring of anomalous methane point sources with multispectral Sentinel-2 satellite

observations, Atmospheric Measurement Techniques, 14, 2771–2785, https://doi.org/10.5194/amt-14-2771-2021, 2021

Jia, M., Li, F., Zhang, Y., Wu, M., Li, Y., Feng, S., Wang, H., Chen, H., Ju, W., Lin, J., Cai, J., Zhang, Y., and Jiang, F.: The Nord Stream pipeline gas leaks released approximately 220,000 tonnes of methane into the atmosphere, Environmental Science and Ecotechnology, 12, 100 210, https://doi.org/https://doi.org/10.1016/j.ese.2022.100210, 2022.