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.

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.

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.

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.
>
> 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.

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'?

The term "standard calibration" might be misconstrued. Perhaps consider an alternative term, such as "naïve calibration"?

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

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.

Fig. 1:

Enhance the caption with more detailed information.

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.

It's worth noting that no background (CH4-free) sea foam pixels are present in the target scene, as depicted in Fig. 1.

The inclusion of the bottom row of Fig. 1 might be redundant since Fig. 3 already encapsulates that information. 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 2 and 3, respectively.

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

Fig. 5:

Consider to add P(Q) in the caption.

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.

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.

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.

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.

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.

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 ...".

l 157:

It might be beneficial to mention why Ueff also varies based on the type of observer, especially for Earth-like imagers.

l 207-208:

You choose 10% because the fraction of negative emissions is roughly 10%?

l 203-204 & l 212:

Are you suggesting that the primary source of uncertainty stems from the uncertainties inherent in the imager's observations?

l 212-214:

A sentence for the conclusion?

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.

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.

l 218:

What does 1M stand for?

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.