In response to feedback from all reviewers, we have substantially revised the treatment of NIMCAM observational uncertainty. Rather than adopting a GOSAT-based estimate, we now include a dedicated uncertainty analysis of the NIMCAM retrieval concept based on radiative-transfer simulations and a detailed instrument uncertainty budget (now detailed in Appendix B). This analysis demonstrates a per-pixel random uncertainty of approximately 40 ppb for background methane concentrations under favourable observing conditions, increasing with solar zenith angle and other observing constraints. All OSSE experiments have been repeated using these revised uncertainty assumptions, and the results throughout the manuscript have been updated accordingly.

Reviewer 1

General Comments:

This manuscript presents a closed-loop observing system simulation experiment (OSSE) to evaluate the capability of a high-resolution multispectral imager (NIMCAM) to constrain diffuse methane emissions over tropical Africa under cloud- and aerosol-limited conditions. The study is well-motivated, the methodology is clearly described, and the results are internally consistent. The work addresses an important limitation of current global satellite observing systems: the loss of clear-sky data in persistently cloudy regions. The results demonstrate that increased spatial resolution can substantially improve clear-sky sampling and, under the assumptions of the OSSE, lead to improved uncertainty reduction in methane flux inversions. This is especially true in regions where lower-resolution instruments provide few to no valid observations, highlighting a robust advantage of the NIMCAM configuration in persistently cloud-covered regions. However, some key assumptions regarding retrieval error structure, sampling behaviour, and transport modelling are not fully characterized within this manuscript. These assumptions are central to the magnitude of the performance gains and may lead to a somewhat optimistic estimate of capability. Some clarification would strengthen the robustness and interpretability of the conclusions.

We thank the reviewer for their positive assessment of the manuscript and for recognising the importance of addressing observational gaps in persistently cloudy regions. We agree that several assumptions in the OSSE framework warrant clearer explanation. We have revised the manuscript to explicitly document these assumptions and to frame our results as representing an idealised, optimistic scenario. Description on the changes made are detailed in response to your specific comments.

Specific Comments:

1: Error scaling and sampling

The NIMCAM spectral configuration (three ~1 nm channels near 1.64 μm), together with the described use of adjacent bands to characterize methane absorption, water absorption, and surface reflectance, appears consistent with a differential or matched-filter retrieval. Such methods can introduce spatially correlated and scene-dependent errors driven by surface reflectance, aerosols, and viewing geometry.

If such correlations are present, the effective sample size is reduced (Neff $\ll N$), and the expected error reduction with increasing sampling density is weaker than $1/\sqrt{N}$. This effect may be particularly relevant for diffuse emission signals, where low signal-to-noise ratios and broad spatial structure increase sensitivity to any correlated retrieval errors.

The authors define super-observation error as:

$\delta = \delta\mathrm{l} + \delta_{\mathrm{GOSAT}}/\sqrt{N}$

Where $\delta\mathrm{l}$ is a lower bound (constant term) and $\delta\mathrm{GOSAT}/\sqrt{N}$ is the random noise reduction. They state that $\delta\mathrm{l}$ represents sub-grid variability and scene-dependent effects.

This appropriately prevents unrealistically small errors as the number of observations increases. However, this term does not explicitly represent reduced effective sample size, spatial correlation, or scene-dependent retrieval errors. As a result, the relationship between sampling density and information content may be sensitive to these assumptions.

The relative improvement of NIMCAM over TROPOMI is evaluated within a consistent OSSE framework that appropriately treats both instruments under the same transport and inversion assumptions. The relative comparison is expected to be robust provided that the underlying error structures are comparable.

However, the NIMCAM measurement approach, based on a small number of narrow spectral channels, may lead to retrieval errors that differ in structure from those of TROPOMI, particularly with respect to spatial correlation and scene dependence. If present, such differences would affect the effective number of independent observations and the scaling of uncertainty with sampling density. As a result, the relative performance gains may be moderately sensitive to assumptions about error independence that do not affect both instruments equally.

The retrieval formulation is not explicitly described. The manuscript does not explicitly define the forward model used to relate measured radiances to methane column enhancements, nor does it characterize the resulting error structure beyond a prescribed per-pixel random error (Section 2.6). The manuscript references Woodwark et al. (2026) for these details, but this information is not included here and Woodwark et al. (2026) is not publicly available at the time of this review (to the best of my knowledge). I understand that Woodwark et al. (2026) may be published before this work. Providing a pre-print during review would help support assessment of these assumptions.

A concise description of the retrieval approach and its expected error characteristics would make the analysis more self-contained and help assess the robustness of the sampling-based performance gains.

We thank the reviewer for this insightful comment. We agree that the assumption of independent errors and $1/\sqrt{N}$ scaling represents an idealisation, particularly for a multispectral retrieval approach where spatially correlated and scene-dependent errors may arise. In this study, we partially account for these effects through the lower-bound term (δl). However, we acknowledge that this term does not explicitly represent spatial error correlation or reduced effective sample size. We also acknowledge there is lacking specific detail on the NIMCAM retrieval approach.

Following these suggestions, we have made the following additions to the manuscript:

- We have clarified that the $1/\sqrt{N}$ scaling represents an upper bound on attainable precision and expanded the description of potential correlated errors arising from surface reflectance, aerosols, and viewing geometry (Added after eq 1. Sec 2.6)
- We have added to the discussion on how deviations from independence would reduce the effective sampling advantage of NIMCAM. (End of Sec 3.4)
- We now include a concise description of the NIMCAM retrieval concept and error characteristics (Sec 2.1 and appendix B).

2: Transport error assumptions

The OSSE uses GEOS-Chem for both the generation of synthetic observations and the inversion (Section 2.6), with differences only in the emissions field. This configuration avoids inconsistencies between forward and inverse modeling but also excludes structural transport error. It is commonly referred to as an “identical twin” OSSE.

In practice, transport uncertainty is a significant component of inversion error and introduces spatially coherent model-data mismatch. While this limitation applies to both NIMCAM and TROPOMI, its impact may not be fully symmetric. The primary advantage of NIMCAM in this study arises from increased sampling density and more spatially distributed observations under cloud and aerosol filtering. In the presence of transport error, additional observations may lead to spatially correlated residuals, which are not explicitly represented in the current error model. The manuscript does note the importance of transport error (Section 3.4), but it is not explored within the current framework. A complementary fraternal twin OSSE would help quantify sensitivity to transport assumptions, although this may be beyond the scope of the current study.

We agree that the use of an identical-twin OSSE excludes structural transport error and therefore represents an optimistic inversion framework. Our intention was to isolate the impact of observational sampling under controlled conditions.

We have revised the manuscript and now:

- Explicitly state in our methods that the OSSE neglects the impact of transport error (Sec 2.6)
- clarify that this likely overestimates inversion performance for both instruments
- Include in the discussion the impact of transport error (Sec 3.4)

A fraternal twin OSSE would indeed provide a more realistic assessment, but is beyond the scope of this study. We do highlight this as a potential direction for future work (Sec 3.4).

Technical Corrections:

This is a very well-written manuscript. I noted only a few minor spelling and grammatical issues:

Spelling errors

Section 2.2

"capabclity" → should be *capability*

"satellities" → should be *satellites*

Section 2.4

"probalistic" → should be *probabilistic*

Minor grammatical / wording issues

Section 1

"taking advantage sensitivity" → should be *taking advantage of the sensitivity*

Section 2.1

"spectral various" → should be *spectral variations*

Section 3.2

"are are primarily influenced" → duplicated word

Section 3.4

"values corresponding to prior knowledge" → slightly awkward phrasing; could be *prior estimates*

"posterior estimate is much closer to the observations" → should likely be *closer to the truth*

We thank the reviewer for identifying these issues and have corrected all these technical errors.

Reviewer 2

This manuscript evaluates the ability of a single satellite, or a constellation of high-resolution satellites (NIMCAM), to improve observational coverage and corresponding inverse estimates over the African tropics using an Observing System Simulation Experiment (OSSE). The study is motivated by the premise that higher spatial resolution enables successful retrievals under partially cloudy conditions where lower-resolution global mappers (e.g., TROPOMI and GOSAT-2) would fail. The authors first quantify observational coverage using synthetic retrievals based on cloud cover and aerosol optical depth thresholds, comparing configurations of 1, 3, and 5 NIMCAM satellites against existing satellite capabilities. They then perform a series of flux inversions using these synthetic observations to assess the impact of increased observational density on emission estimates.
This manuscript addresses an important question regarding the satellite capabilities needed to improve methane flux estimates in the African tropics. The work is well- structured and clearly presented. I recommend acceptance pending minor revisions.

We thank the reviewer for their positive assessment of our work and appreciate the recognition of the importance of this question and the clarity of the manuscript. We have carefully addressed the minor revisions suggested below to further strengthen the manuscript.

My primary concern relates to the treatment and use of super observations and their associated error assumptions in the OSSE:

- The authors assume GOSAT-like observational uncertainty for NIMCAM. Can they justify whether this assumption is appropriate given the differences in spectral resolution and retrievals between the instruments?
  We thank the reviewer for raising this important point. The use of GOSAT-like uncertainty is a simplified estimate of NIMCAM retrieval noise. We have now revised our work to instead use a detailed NIMCAM-specific derived uncertainty (detailed in appendix B). All results are updated accordingly.

- In Table 1, the authors assume that super observation uncertainties decrease according to the central limit theorem (CLT). However, Equation (1) appears to impose a lower bound due to correlated error. In practice, super observations are likely to retain some degree of correlated bias and may not strictly follow CLT scaling. The analysis would be strengthened by including sensitivity tests that account for varying levels of correlated error in the super observations.
  We thank the reviewer for raising this. The limitations of the CLT assumption were also addressed in response to reviewer 1 and recognition of the assumptions made are now included in the manuscript discussion (Sec 3.4). While a full sensitivity analysis exploring different levels of correlated error would be valuable, we consider this beyond the scope of the present study but do highlight it as an important direction for future work (in conclusion, Sec 4).

- How are super observations aggregated (e.g., by orbit, by day, or over a 10-day period)? Typically, aggregation is performed at the orbit scale to reduce random error. Aggregating over a 10-day window would potentially cause excessive temporal smoothing.
  We aggregate the super observations over a 10-day period.~~,~~ driven by the 10-day repeat cycle of the narrower NIMCAM orbit track. We have now clarified this definition of a super observation in methods, Sec 2.5.

  We also note the limitation of temporal smoothing on results, which may reduce sensitivity to short-timescale variability. However, for the specific application in this work, diffuse methane emissions from tropical wetlands typically vary on timescales longer than days~~,~~ and are driven by basin-scale hydrological processes. As a result, aggregating observations over a 10-day period is expected to provide a reasonable representation of these emissions while allowing for consistent spatial sampling across the NIMCAM orbit cycle.

- Is the observational error covariance matrix assumed to be diagonal? Does it use a lower limit for sigma like in Equation 1, or does it follow the CLT? Additionally, what observational error covariance (S_o) is used for the TROPOMI inversions? Are super observations also applied to the TROPOMI data?

  The error covariance matrix is assumed to be diagonal for the super observations, partially because we have limited information about spatial correlations in the gridded data. Also the number of gridded observations is relatively small, we do not impose a lower limit on the aggregated observation error described by the error covariances. We process the TROPOMI data following the same approach used in NIMCAM to construct the super observations, except that we use the actual individual observation errors and the number of observations within each grid cell to compute the random component of the observation error.

- Please also clarify how retrievals that span multiple grid cell boundaries are handled.
  Retrievals are assigned to GEOS-Chem grid cells based on their centre location, such that each retrieval contributes to a single grid cell. Given the fine spatial resolution of NIMCAM relative to the model grid, this approximation introduces only a small error and is unlikely to affect the results. We thank the reviewer for raising this point and have now described this in the methods, Sec 2.5.

- In the discussion of super observations, it would be appropriate to cite foundational work such as Eskes et al. (2003), or Chen et al. (2023) for a CH_4-specific application.
  Thank you for sharing these relevant studies, we have now cited these in the description of super observations, Sec 2.5.

Specific comments:

- Figure 3: For clarity, consider adding vertical dashed lines to delineate monthly boundaries, or include an x-axis label in the top panel.
  This figure has been updated.
- Figure 7b: Please clarify whether this panel shows percentage or fractional error reduction.
  This is fractional error reduction, and units are now stated explicitly.
- Line 171: Please confirm that true boundary conditions are prescribed in the prior simulation to avoid bias from boundary effects.
  Yes they have the same boundary conditions, and this been included as a detail in our methodology.
- Line 173: The standard term is "Observing System Simulation Experiment (OSSE)"; consider using this instead of "closed loop system."
  We have changed the naming 'closed-loop system' to the standard term throughout the manuscript.

Technical Corrections:

- Line 98: Typo in "capability"
- Line 136: Typo in "probabilistic"
- Line 202: Typo in the subscript of Sobs
- Line 275: Typo ("O" at end of line)

We thank the reviewer for identifying these issues. All technical errors are now fixed.

Reviewer 3

This study evaluates the potential for NIMCAM to constrain diffuse methane emissions in Africa. NIMCAM has a unique instrument design: 60 m pixels, no targeting, and only 3 (effective) spectral bands, on and off an absorption feature of methane ~1640 nm. The authors argue that, for typical satellite retrieval quality filtering, NIMCAM offers expanded observational coverage (relative to in particular TROPOMI) via its small footprint not being as impacted by clouds and aerosols. Through OSSEs, the authors further argue that NIMCAM can outperform TROPOMI for flux inversions in select months. The point that higher-resolution observations are valuable for seeing through the clouds in the tropics is well taken, but I do not believe the authors reach the bar of showing that NIMCAM would be useful for constraining diffuse emissions on the continental scale (let alone more useful than TROPOMI). I believe that the paper is not suitable for publication in its current form.

We thank the reviewer for their careful and constructive assessment. We emphasise that it was not our intention to imply that NIMCAM is more useful than TROPOMI; rather, our aim was to assess the potential advantages of high-resolution observations for improving sampling in regions affected by clouds and aerosols. We have carefully addressed all the suggested revisions and comments as follows.

General comments

As I understand it, the authors use fluxes from Feng et al. (2023) that are at 2 x 2.5 degree resolution, optimized using GOSAT. They regrid these fluxes to 0.25 x 0.3125 degree resolution, run them through GEOS-Chem, and then sample the simulated atmosphere using NIMCAM or TROPOMI observation locations. They then ingest these pseudo observations (with added noise) in an ensemble Kalman filter using the same prior emissions that GOSAT started with and see how well NIMCAM or TROPOMI can recover the true fluxes. I have the following concerns:

- In general, I was not at all impressed by the OSSE results. Maybe this is related to the errors detailed below. In Figures 7 and 8, comparing “Prior-True” and “Post-True” on the fluxes, it seems that none of the instruments do particularly well, except for NIMCAM correcting some of the overestimated fluxes in the Congo. A lot of the error reductions are in areas where the proximity to the truth actually gets worse. Also, hasn’t this same research group constrained tropical African methane fluxes using TROPOMI (Lunt et al., 2021, https://doi.org/10.1088/1748-9326/abd8fa), for example the Sudd, but now they are saying TROPOMI is useless for this purpose?

  We are not sure what areas the reviewer is concerned about where proximity to truth gets worse. In all regions we presented an improvement in the observation model agreement in the posterior fluxes compared to the prior (lighter colours in the Post-True map than the Prior-True map).

  Note that the OSSE experiment results are now updated with the more robust NIMCAM uncertainty choices. We have included quantified results on the relative improvements to the truth, as well as total error reduction, to give a clear comparison between the different NIMCAM constellation systems as well as TROPOMI.

  TROPOMI is certainly already a useful dataset for constraining methane fluxes, including in tropical regions. What we hope to demonstrate in this paper is that higher resolution observational systems such as NIMCAM, can allow for further gain of information due to improved data coverage particularly in cloudy regions. We hope to demonstrate that there are more significant gaps in the TROPOMI observation set over the Congo basin, relative to NIMCAM observations (in the 1, 3 and 5 satellite orbit regimes).

  However, none of this is to say that TROPOMI is useless for the purpose of constraining diffuse emissions in this region. The OSSE results clearly show that TROPOMI data does improve proximity to the truth and shows widespread error reduction across the region. We would argue that a higher resolution dataset such as NIMCAM would just offer complementary information.

- Is there any proof that NIMCAM will be able to provide an accurate but low precision methane retrieval? There is only a citation to a paper that is not accessible (Woodwark et al., 2026). Have the authors shown that a 3 spectral channel retrieval can avoid systematic biases that would preclude its use in an inversion (perhaps this could be done by coarsening the spectral resolution of existing spectra covering ~1640 nm, such as MethaneSat?)

  We thank the reviewer for raising this important point. The NIMCAM retrieval has been designed and tested to exhibit a strong linear response to variations in methane, which is the primary requirement for its use within an inversion framework. While the retrieval does exhibit an absolute bias in retrieved columns, this bias is largely consistent across scenes and can in principle be accounted for through calibration or bias correction.

  We agree, however, that the more critical concern is the presence of spatially variable systematic biases arising from radiative transfer effects, such as aerosols, thin cloud, and surface reflectance variability. These effects can introduce scene-dependent errors that do not average out and may be misinterpreted as real methane variability in an inversion.

  To address this, we have extended the discussion in the manuscript to include the use of a CO_2 proxy retrieval approach. This involves incorporating an additional spectral channel sensitive to CO_2 absorption and normalising the methane retrieval with the corresponding CO_2 retrieval. Because atmospheric CO_2 is well mixed at the spatial scales considered, spatial variations in the CO_2 signal primarily reflect radiative transfer effects. Taking the ratio of CH_4 and CO_2 therefore suppresses these shared artefacts, reducing spatially correlated bias.

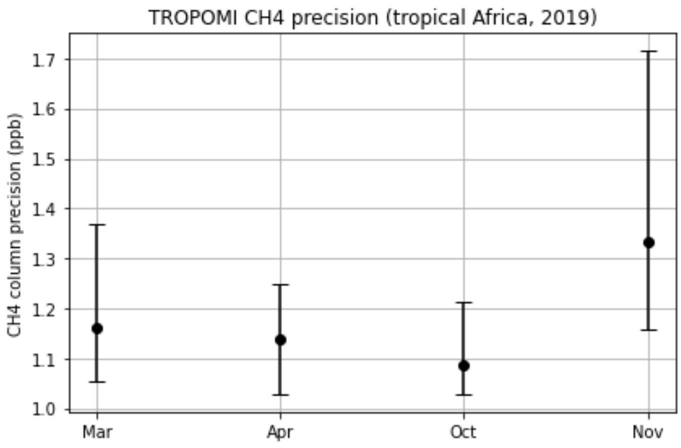
  A description of this approach has now been added in Section 2.1 as a realistic pathway to mitigate systematic biases in a NIMCAM-type retrieval.

- Do you add 15–30 ppb of noise to both TROPOMI and NIMCAM data before ingesting them? Is this correct given the estimated precisions of each instrument? Why is this different than the 2.5e15 and 2.5e17 molecules/cm2 per-pixel precisions given later on? By the way, using 2e25 molecules dry air/cm2, these translate to 0.125 and 12.5 ppb, which are overly optimistic. TROPOMI has a precision ~12.5 ppb (cf. Schneising et al., 2026, https://doi.org/10.5194/amt-19-2407-2026), a factor of 100 difference, so I encourage the authors to clarify how these precisions were determined.

  The uncertainty treatment has now been substantially revised and clarified throughout the manuscript.

  The 15-30 ppb has been replaced by an instrument specific uncertainty model derived from a dedicated NIMCAM concept uncertainty analysis (appendix B). This analysis indicates a per-pixel random uncertainty of approximately 40 ppb under favourable observing conditions, increasing to ~80 ppb at high solar zenith angles. The OSSE therefore now uses a solar-zenith-angle-dependent NIMCAM observation uncertainty rather than a fixed value. All uncertainty values are now reported in ppb for consistency.

  The previously reported 2.5E15 molec/cm2 precision value for TROPOMI was derived from the TROPOMI level 2 precision dataset over the region and dates range of the study. However, we identified an error in our processing that caused this value to be underestimated. Following correction, the mean Level 2 TROPOMI precision is approximately 1.1–1.3 ppb, depending on the month (see plot below).



We note that the reported TROPOMI level 2 data are lower than values of ~10–15 ppb reported in the literature (e.g. Schneising et al., 2026). This difference arises because the Level 2 precision values represent internal retrieval uncertainty estimates based primarily on measurement noise, whereas validation-based estimates (e.g. from comparisons with TCCON) include additional sources of systematic error. The uncertainty values used in this study are intended to represent the random component of observational error (e.g. instrument noise and retrieval precision). Systematic biases are treated separately, as they do not decrease through averaging in the same manner as random errors when constructing super observations. Within the OSSE framework, the Level 2 precision values are therefore used to characterise the random observational uncertainty, while residual systematic effects are represented through the lower-bound observational error term (δl = 5 ppb).

Further to explain the 2.5E17 molec/cm2 value for NIMCAM, this was calculated as the single pixel precision value derived from an earlier radiative transfer retrieval concept. This has been superseded to the more comprehensive uncertainty analysis, with the revised 40-80 ppb range.

- Can the authors confirm that the square root should be applied in this manner in Equation 4?

  This is consistent with a standard uncertainty reduction calculation used in inversion studies.

- Line 169: what do you mean when you say you assume the prior fluxes are the same as Feng et al. (2023)? Can this not be checked?

  We use the same prior inventories as Feng et al. (2023). CH4 has many different sources, and we now list the most important parts.

- What is the rationale for Equation 2? Why not use the same prior uncertainty as Feng et al. (2023)?

  The prior uncertainty is prescribed to support the aims of the OSSE experiment. Its magnitude and spatial distribution are chosen to permit meaningful updates to the prior fluxes during the inversion, allowing the analysis to focus on the influence of observational coverage and measurement uncertainty rather than being constrained by an overly restrictive prior.

- How is the number of clear sky retrieval determined (i.e., how do you align spatially the 60 m S2 cloud probabilities with 1 km AOD that has been downscaled to, I believe, 1 km probabilities of 60 m pixels being lower than a threshold). I believe the sentence starting at Line 190 is incomplete?

We thank the reviewer for this comment and will clarify this in more detail in the revised manuscript. The reviewer is correct that we only have aerosol optical depth (AOD) information at 1 km resolution, expressed as the probability of exceeding a given threshold, and this cannot be directly aligned with the 60 m Sentinel-2 cloud probabilities at the sub-pixel scale.

To address this, we assume that AOD conditions are randomly distributed within each 1 km pixel. Under this assumption, the probability of a given 60 m sub-pixel being below the AOD threshold is uniform across the grid cell. The number of clear-sky pixels within each 1 km box is therefore estimated as the number of cloud-free 60 m pixels multiplied by the probability that AOD is below the threshold (i.e. P(AOD<threshold)).

We have added detail in Section 2.5 to describe this calculation in more detail. Sentence on line 190 has also been adjusted.

- What value is used for the lower limit of uncertainty in Equation 1?

  5 ppb, this is now stated explicitly

- I was confused by Line 340 since your OSSE does not vary transport.

  Though we did not vary transport in the OSSE, there is still an assumed transport error in the model to account for their impacts on posterior flux estimates. This will be described more carefully.

Minor comments

- The authors refer to an instrument with 1 km spatial resolution and a 5-day repeat as "nominally GOSAT-2." I believe GOSAT-2 actually has 10 km spatial resolution with a 6-day repeat (cf. Barr et al., 2025, https://doi.org/10.5194/amt-18-6093-2025).
  We thank the reviewer for noting this clear mistake in comparing this instrument to GOSAT-2. We have removed this comparison, and now just refer to this as a notional mid-resolution instrument.

- The source of the 60 m AOD data used in Appendix B is not stated.
  This is an AOD product derived from Sentinel 2 imagery data, we have included this description in the appendix. We have also added an acknowledgement to Feng Ying (UCL) for proving this data for the study.

Specific comments

- Line 7: "2–5" should be "3 and 5"
- Line 35: typo around "sensitivity"
- Line 48: remove comma after "region"
- Line 77: "various" to "variability" (?)
- Line 139: 0.315 to 0.3125
- Line 154: add W on 20 degrees
- Line 202: "obs" isn't subscripted

We thank the reviewer for identifying these issues. All technical errors are now fixed.