

Reviewer 1

Specific Comments

Comment: Figure 3.3d – page 10 – the caption says this is a “representation”. Is this real data that are shown, or some kind of conceptual illustration? I suspect it's the former. If so, please add which date/flight this was to the caption and make clear it “represents” real data.

Response: It is from a real survey. We have updated the text to: "Three-dimensional visualisation of measured downwind CH₄ enhancements from a flight performed by the in-situ aircraft on 28 September 2023."

Comment: Line 285 and figure 4 – I appreciate the text says that the MAMPA2D emissions method is described in detail elsewhere, but I would recommend adding a brief summary of it here such that the figures and narrative can stand alone better in this paper. What is the column anomaly relative to? I.e. what is the background column (and column units) that the percentages in figure 4 deviate from?

Response: We agree and added some information provided in SI-5 to the main text:

L282-287 reads: “In short, downwind of a source, multiple cross-sections were explicitly defined (i.e. geometrically constructed) through the observed plume, spaced at 10 m intervals. Within each cross-section, CH₄ was sampled at 10 m resolution. The transported mass of CH₄ per unit time (t h^{-1}) was then calculated for each cross-section using the component of wind speed perpendicular to the transect. This process also included a dedicated normalisation step for each cross-section, in which the plume was separated from the surrounding background on either side. This allowed the local CH₄ background to be estimated and subtracted prior to flux calculation.”

The column anomaly is relative to the absolute atmospheric CH₄ background column (in molecules per cm²) given for a specific ground scene. As indicated in the supplementary SI-3, we use dedicated simulated radiances for each flight day, which reflects, e.g., the atmospheric conditions during that day but also

geometrical parameters like surface elevation. The CH₄ background column used in those radiative transfer simulations is extracted from SLIM model [Noel et al., 2022]. Furthermore, we use a look-up table approach concerning surface elevation so that the background column of each ground scene is modified accordingly (higher surface elevation -> shorter column -> lower background methane dry column mixing ratio).

For the cross-sectional flux method (see also SI-5, Eq. S2) to compute the mass of CH₄ advected, we use the amount of the CH₄ dry column or mole fraction anomaly in units of molec cm⁻², which is the dry column anomaly multiplied by atmospheric background dry column amount of the respective ground scene.

Comment: Line 318 – Hysplit has been used as a guide to the footprint of sources. I think this is fair enough, but it's not the best air history model around and relies on some sketchy meteorological inputs available compared to e.g. WRF-STILT or FLEXPART, or NAME. I'm not suggesting those models must be used here as this seems like a quick and ready sense check. But did you check for any wild divergence in footprint if you try trajectories further up or down in the mixing layer to check consistency of the assumed footprint, or to constrain any potential uncertainty?

Response: We thank the reviewer for this thoughtful comment regarding the limitations of HYSPLIT relative to higher-resolution Lagrangian models such as WRF-STILT, FLEXPART, and NAME, and the potential sensitivity of footprints to release height within the mixing layer. We agree that this warrants testing.

To address this, we performed a sensitivity analysis designed to test whether HYSPLIT footprint uncertainty could affect our quantitative conclusions. We selected three representative curtain quantifications spanning the minimum (Curtain 230928_C1), mean (Curtain 230917_C3), and maximum (Curtain 230926_C1) other-source contributions observed in our dataset, where percentage contributions refer to the CH₄ source contribution relative to the total curtain emission rate. For each case, we re-ran HYSPLIT trajectories at 10 m AGL and at the top of the mixing layer, bracketing the midpoint release height used in the main analysis. By choosing cases across the full range of other-source contributions, we tested whether release-height sensitivity could matter most in

the maximum-contribution scenario, where footprint sensitivity would be expected to have the largest influence on our results.

The minimum and mean cases were essentially insensitive to release height (range ≤ 0.2 percentage points; Table S12). The maximum-contribution case showed a larger range of 4.5 percentage points (8.1 to 12.6%), with the base case representing the upper end of this range. While this indicates some release-height sensitivity in cases with substantial other-source contributions, the qualitative source-region identification remained consistent across all release heights, and the quantitative variation is within the overall uncertainty bounds of our curtain quantification methodology.

We have added a brief description of this analysis to the results section of the revised manuscript (lines 908-911), with the full sensitivity analysis and supporting table presented in Section SI-11 (Table S12) of the supplementary material.

L908-911 reads: "A sensitivity analysis of HYSPLIT footprints to release height, performed for cases spanning the minimum, mean, and maximum other-source contributions (relative to the total curtain emission rate) in our dataset, confirmed that our source-region identification and quantitative estimates are robust to release-height uncertainty (see SI-11)."

Comment: Table 1 – is really long.... Can you summarise and highlight/colour/bold some of the more salient results you want readers to take from it? And, again, the caption is very long – move some of this to the methods narrative as it contains lots of info on data filtering that should not be in a caption.

Response: We have made several changes to streamline the table. The totals have been removed, as they are already presented in Table 3, reducing the table by five rows. Consequently, references to captions "e" and "g" have also been removed from the figure caption. Caption "e" has now been transferred to Table 3, where it is more appropriate. We have also condensed the table title and caption significantly, instead referring the reader to the text and SI for further details.

Comment: Figure 5 – caption does not say what the solid black line is? Guessing this is a line of best fit to all data? Least square fit?

Response: The solid black line is the 1:1 line, which is included because the figure compares two estimation methods. We have now explicitly labelled it as the "1:1 line". Since the raw data in this figure have not been separated according to the underlying reporting methods (e.g. Methods 1, 2, and 4), no regression lines are presented here (or in Fig. 6). It is only after the data are separated into their respective cohesive groups (Groups 1–6) that regression analyses are appropriate. We have also clarified the presence of the 1:1 line in Figs. 6 and 8.

Technical Comments:

Comment: Line 167 – space needed between quantity and unit (Mt). Several other instances to check.

Response: We have added spaces here and in other instances where this occurred.

Comment: Figure 2 caption – this is a very long and descriptive caption. Can some of it be moved to the narrative (esp. for what is not shown in the figure)?

Response: We have reduced the caption length by 130 words and now refer the reader to SI-2 for further details. We also shortened the caption for Figure 4, as it contained unnecessary detail for a figure caption.

Comment: Figure 2 right panel – only coal mines register on the y-scale. Would a log-10 y axis help to resolve the other sources better?

Response: We agree and have updated the y-axis to a log10 scale.

Reviewer 2

Specific comments:

Comment: Line 67: It sounds like solely facilities that emit more than 100,000 t CO₂-e per year report under the Safeguard Mechanism. However, in SI-1 the table shows the Safeguard Mechanism database listing also coal mines that emit less. Make clear, that the Safeguard Mechanism is a database and what reporting threshold it sets.

Response: We agree that the wording could be clearer and have revised the text accordingly.

The Safeguard Mechanism is a regulatory framework rather than a database; emissions data are reported under the National Greenhouse and Energy Reporting Scheme. The Safeguard Mechanism applies to facilities with total Scope 1 emissions exceeding 100,000 t CO₂-e yr⁻¹, where CO₂-e aggregates emissions across multiple greenhouse gases (CO₂, CH₄ and N₂O).

The dataset may include disaggregated emissions (e.g. methane from coal mines) that are individually below this threshold, while the facility's total CO₂-e emissions exceed 100,000 t yr⁻¹. This distinction explains the presence of coal mines with lower reported CH₄ emissions in Table SI-1.

We have revised the manuscript to clarify both the reporting threshold and that it applies to total CO₂-e emissions aggregated across gases, rather than to individual gases or emission sources. We have also added an additional reference to link directly to the Clean Energy regulator website describing the framework (CER, 2026).

L67-69 now read: "Facilities with total Scope 1 emissions exceeding 100,000 t CO₂-e a⁻¹ (aggregated across CO₂, CH₄, and nitrous oxide (N₂O)) are covered by the Safeguard Mechanism, a regulatory framework which requires emissions to remain below a declining baseline."

Comment: Line 156: Up to this line it is not yet clear what is meant by applying two different approaches to estimate emissions. Here, "across aircraft" suggests that two aircraft were used. Make it clearer at an earlier stage that the two approaches are in-situ and remote sensing (e.g. Line 30 in the Abstract).

Response: We agree that this should be clarified. The sentence has been revised to: "The aim of this study was to use an in-situ and a remote sensing aircraft-based approach to quantify CH₄ emissions [...]". We have also explicitly stated that two aircraft were used by adding the phrase: "[...] and across the two aircraft". In addition, we have updated the abstract to improve clarity: "Here, an in-situ and a remote sensing aircraft-based approach were used to quantify CH₄ emission rates from 17 coal mines in the Bowen Basin, Australia."

Comment: Line 411 and Table 1: Are the calculated emission rates compared to Method 4 reporting (without distinguishing between CEM and PEM) or to the sum of Method 4 and Method 1 reporting?

Response: For all underground facilities reporting methane emissions under the NGER scheme, emissions are reported at the facility level in aggregated form, with no publicly available breakdown by method. Therefore, where facilities report the use of both Method 1 (or Method 2) and Method 4, the aircraft-derived estimates are compared against total reported facility emissions (i.e. the combined contribution of these methods), rather than individual methods.

Thus, it is not possible to disaggregate and assess individual NGER methods (or their methods of operation) for underground coal mining, and these are instead assessed collectively on an aggregated basis. We have revised the manuscript to clarify this point.

Furthermore, upon re-examination of the underlying dataset and its documentation of methods used by operators to estimate emissions, we clarify that this dataset refers specifically to methods used for determining fugitive CH₄ emissions only. We have therefore removed the statement suggesting uncertainty in gas-specific method attribution, as this is incorrect. The remaining reporting limitations (i.e. lack of disclosure of CEM/PEM for Method 4 and facility-level aggregation with no breakdown by method) remain valid and are retained in the revised manuscript.

L402-407 now read: "However, the use of CEM or PEM to derive Method 4 estimates is not publicly disclosed under the Safeguard Mechanism. In addition, CH₄ emissions are aggregated at the facility level, with no publicly available breakdown by method, meaning comparisons were made against total reported facility emissions (i.e., the combined contribution of Method 1/2 and Method 4).

As a result, individual NGER methods (and their methods of operation) for underground coal mining could not be separately assessed in this study and were instead evaluated on an aggregated basis.”

Comment: Line 566: Please mention where you take the coal production data from.

Response: We have added the reference and included a citation to SI-1, which presents all production data. For further clarity, we have also added this reference to the Fig. 6 caption (L452).

Technical corrections:

Comment: Line 138: better use “helicopter-borne measurement system” instead of “helicopter-borne aircraft”

Response: This has been updated.

Comment: Line 180: better use FY24

Response: This has been updated.

Comment: Figure 2 A: For the orange “field” in the North, the shape of the markers is not visible. It is not clear whether there are only gas development wells or also CMWG power stations or ammonium nitrate plants present. Maybe show some markers, e.g. gas development wells, with an outline.

Response: We have improved the visibility of marker shapes by adding dark outlines to filled markers, so individual symbols remain distinguishable even in dense clusters. We have also added an inset panel showing a magnified view of the northern cluster, where the gas development wells, CMWG power stations, and ammonium nitrate plant are now clearly identifiable as distinct shapes.

Comment: Figure 2 B: Emissions from all other sources than coal mines are barely visible. You could include a second axis at a reduced scale.

Response: In line with the first reviewers suggestion, we have updated the y-axis to a log10 scale so that the emission strengths of other sources are visible.

Comment: Line 580: “between the two approaches” instead of “between the aircraft”

Response: This has been updated.

Comment: Line 858: typo: aggregate

Response: We note that line 858 spells 'aggregate' correctly, so have made no changes.

Author additional corrections (editorial)

In addition to the revisions made in response to reviewer comments, the authors identified several further corrections and editorial improvements during manuscript revision. These changes are summarised below.

Correction 1

During review, we noted that the production values for MC-1 were transcribed incorrectly. Updated production data for MC-1 were incorporated, resulting in only minor changes to the total values in the document:

- Total ROM coal extraction increased slightly from 254 Mt to 256 Mt.
- Surface mining total remained unchanged in absolute terms (215 Mt) but decreased slightly in proportional share (85% to 84%).
- Underground total increased slightly (39 Mt to 41 Mt), with a corresponding small increase in share (15% to 16%).
- Within breakdowns, open-cut remained unchanged (198 Mt; 78% to 77%), underground subcategory remained unchanged (27 Mt; 11% to 10%), and mine complexes increased slightly (29 Mt to 31 Mt; 11% to 12%).

Figure 6 has been updated accordingly, as it incorporates the percentage of underground ROM production, which has now changed, with no change to conclusions. This also led to small changes in the emission intensities (EIs) reported in the text and Table 4 for Groups 1–6. This correction does not affect the interpretation of the results and conclusions presented in the manuscript.

Correction 2

During revision we noted that the average EIs in Table 4 had been calculated using aggregated totals rather than the appropriate mean values. These have now been recalculated using the mean values reported in Table 1. This correction does not affect the interpretation of the results and conclusions presented in the manuscript. The mean EI based on aircraft quantifications for surface mines was revised from $4.4 \pm 4.2 \text{ kg CH}_4 \text{ t}^{-1} \text{ ROM}$ to $5.6 \pm 4.2 \text{ kg CH}_4 \text{ t}^{-1} \text{ ROM}$, and for underground mines from $5.3 \pm 2.7 \text{ kg CH}_4 \text{ t}^{-1} \text{ ROM}$ to $5.4 \pm 2.7 \text{ kg CH}_4 \text{ t}^{-1} \text{ ROM}$. The conclusion that mean EIs are similar between groups remains unchanged.

Correction 3

In addition to the revisions made in response to reviewer comments, several editorial changes were made to improve the precision and scientific framing of the manuscript.

Abstract (L38–40): The statement *"These results demonstrate for Australian coal mines, direct monitoring approaches to quantify underground mine emissions are fit for purpose, but bottom-up surface mine emission estimation methods require review."* was revised to *"Alignment between airborne top-down and measured vented underground coal mine emissions increases confidence in Australia's Tier 3 direct monitoring methods, while the difference between top-down and bottom-up surface mine emission estimates suggests core-based methods require review."* This change avoids the non-scientific phrase "fit for purpose" and reframes the conclusion in terms of the observed agreement between independent estimation approaches and the confidence this provides.

Discussion (L670–672): The statement *"By these measures, the results provide statistically supported evidence that the predominantly measurement-based inventory method used to report fugitive emissions from underground coal mines (NGER Method 4) in the Bowen Basin are fit for purpose."* was revised to *"Collectively, these results provide statistically supported evidence of agreement between airborne estimates and operator-reported emissions derived using NGER Method 4, increasing confidence in both the airborne and inventory-based approaches."* This change replaces the subjective phrase "fit for purpose" with a more precise description of the observed agreement between independent estimation approaches and its implications.

Discussion (L719–722): The statement *"By contrast, Tier 3 direct measurement approaches for underground mines, as demonstrated here for Australia and supported by United States studies using Carbon Mapper aircraft-based remote sensing surveys (Penn, 2025) and Polish studies using HELiPOD (Förster et al., 2025), appear to provide reliable estimates suitable for underpinning credible national reporting and regulatory compliance."* was revised to *"By contrast, Tier 3 direct measurement approaches for underground mines, as demonstrated here for Australia and supported by studies in the United States (Penn, 2025) and Poland (Förster et al., 2025), show consistent agreement between measurement-based estimates and inventory-based reporting frameworks across regions."* This change removes

performance- and policy-oriented language and reframes the statement to focus on consistency evidenced across independent measurement studies.

Conclusion (L848–851): The statement *"When compared with bottom-up estimates at both the facility- and aggregate-levels, airborne estimates from underground mines showed non-significant bias and good agreement with mean annual emission rates, suggesting that Australia's Tier 3 direct monitoring approaches for underground mines in this basin have reliable performance."* was revised to *"When compared with bottom-up estimates at both the facility- and aggregate-levels, airborne estimates from underground mines showed non-significant bias and good agreement with reported mean annual emission rates. This agreement increases confidence in the estimates produced by both methods and supports Australia's current Tier 3 direct monitoring approaches for underground mines."* This change replaces a direct performance assessment with a more evidence-based interpretation of the observed agreement between independent estimation methods and its implications for confidence in the resulting emissions estimates.

Global context statement (L868–871): The statement *"Within a global context, findings from the Bowen Basin highlight that IPCC Tier 2 and 3 inventory methods for surface mining require careful implementation and independent verification, while also adding to existing evidence that Tier 3 measurement-based approaches for underground operations provide reliable estimates suitable for national reporting."* was revised to *"Within a global context, findings from the Bowen Basin highlight that IPCC Tier 2 and 3 inventory methods for surface mining require careful implementation and independent verification, while the observed agreement for underground operations is consistent with existing evidence from measurement-based Tier 3 approaches applied in underground mines."* This revision removes evaluative language regarding method performance and reframes the statement in descriptive terms of observed differences and consistency with existing evidence.

Correction 4

Other minor edits to wording include:

L113: The word "applied" was changed to "deployed" to improve consistency in methodological terminology.

L115: The phrase "These aircraft" was revised to "The measurements made by these aircraft" to avoid ambiguity regarding attribution.

L125: "can" was removed.

L675: The phrase "For underground coal mines (Group 1)" was revised to "For the case of underground coal mines (Group 1)" to improve clarity.

L762: The word "those" was added in "such as those from cattle" to improve clarity.

DCCEEW Comments

Comment: In the study, uncertainties are alternately presented as 1 or 2 σ . Is there a reason for this? We note Borchardt et al. 2025 (e.g. in the caption to Table 1) was presented using 2 σ , which is the more standard practice.

Response: While both 1 σ and 2 σ uncertainty representations are used in the literature, we agree that presenting uncertainties consistently using 2 σ (approximately 95% confidence intervals) improves comparability with prior studies, and we will revise the manuscript accordingly. We have updated this throughout the text and in the relevant figures (Figs. 5, 6, and 7).

Comment: Suggest modifying line 804, to avoid the possible implication that intra-annual emissions data is collected but not disclosed under NGER. NGER data is both collected and disclosed (for facilities covered by the Safeguard Mechanism) on an annual basis.

Response: We agree that this statement could be clarified. We will revise the wording to make this distinction explicit and avoid any implication that sub-annual data are collected but not disclosed by the regulator.

We have revised the sentence to:

“Higher temporal resolution emission data (e.g., monthly, weekly, or daily) may be used by operators under Method 2 and Method 4 to derive annual estimates; however, under the NGER scheme, emissions are collected and reported on an annual basis.”

Comment: On line 531, results of the statistical significance test comparing reported and observed emissions are presented for all open cut mines excluding those mines studied in Borchardt et al. 2025. If the Borchardt et al. 2025 data is instead excluded from the Method 2 group (Group 4, line 541), is the difference in slope still statistically significant (noting the small sample size and limited emissions range)?

Response: We initially did not present this because excluding the mine studied in Borchardt et al. (2025) leaves only three data points, which limits the statistical power of any analysis. For completeness, exclusion of this mine yields the following metrics for the open-cut (Method 2) group:

- Mean bias decreases from 2.8 t hr⁻¹ to 0.6 t hr⁻¹ (p = 0.09, not statistically significant at the 0.05 level)
- Slope increases from 2.4 to 3.4 [1.7, 5.0], p = 0.03 (statistically significant)

- NRMSE decreases slightly from 0.84 to 0.80

On an aggregated basis, the measured total is 2.4 t hr^{-1} vs reported total of 0.65 t hr^{-1} , a ratio of 3.8. This is an increase from 2.5 in the full dataset.

While the slope remains statistically significant, the mean bias is not, due to the small sample size. Nevertheless, the overall conclusion of a positive bias in Method 2 emissions remains consistent, as the discrepancy between aggregated totals increases.

No changes were made to the manuscript.

Comment: Following from 3, is the statement on line 858 correct that comparisons with both “current Tier 2 and Tier 3” (emphasis added) methods show statistically significant positive bias? It’s not clear that this sentence and line 542 are consistent.

Response: We agree that this statement requires clarification. When assessed separately, only surfaces mine applying Method 1 show statistically significant positive bias at the facility level, while Method 2 mines do not reach statistical significance due to limited sample size. However, when considered in aggregate, the surface mine dataset exhibits a statistically significant positive bias.

We will revise the wording in line 858 to distinguish between aggregate-level significance and the more limited statistical power of individual subsets, ensuring consistency with line 542.

We have updated the text to the following:

“When considering all open-cut mines combined, measured emissions were 3.6 times bottom-up totals, increasing to 5.8 times for Method 1 alone. For Method 2, aggregate measured emissions were approximately 2.5 times bottom-up totals. Facility-level results also indicated a consistent positive bias, although not all groups reached statistical significance, with Method 2 in particular not achieving statistical significance. This is likely a consequence of limited sample size, rather than providing evidence of no bias.”

Comment: Results presented in Figure 6 and Table 4 suggest that “the intensity of CH₄ emitted per tonne of coal [extracted is] independent of mine type” (line 571). If this is true, does it imply that the Langmuir gas content model for coal, presented in e.g. Kholod et al. 2020, is not applicable to understanding emissions from coal extraction? Isothermal gas content is generally assumed to increase with depth for all coal ranks because of increasing pressure. Is there a reason why

coal seams at open cut mines would be relatively oversaturated with gas compared to underground mines?

Response: We agree that our findings do not imply that the Langmuir gas content model is inapplicable to coal mine CH₄ emissions. We have revised the discussion and conclusion to clarify this point.

Specifically, we now explain that the observed similarity in emissions intensities between open-cut and underground mines likely reflects overlapping depth and gas-content ranges between mine types, rather than a breakdown of Langmuir-type gas adsorption behaviour. While underground mines are often assumed to have higher gas contents due to greater depth and pressure, many modern open-cut operations in the Bowen Basin extend to substantial depths where coal seams may also retain significant gas content.

We also clarified that mixed thermogenic isotopic signatures observed in available isotope samples may indicate additional microbial CH₄ contributions at some mines.

Accordingly, we revised the text to avoid implying that the results contradict the underlying gas adsorption model, and instead emphasise that mine type alone may not be a reliable proxy for CH₄ emissions intensity.

The text now reads: “Despite the high variability among open-cut EIs, the comparability between underground and open-cut mines suggests that mine type alone may not be a reliable proxy for CH₄ EI. While underground mines are conventionally assumed to emit more CH₄ due to greater depths and associated gas contents (IPCC, 2006), many modern open-cut operations extend to substantial depths where coal seams may retain significant gas content consistent with Langmuir-type gas adsorption behaviour (e.g., Kholod et al., 2020). The similarity in EI observed here therefore likely reflects overlapping depth and gas-content ranges between mine types, rather than a breakdown of the underlying gas adsorption model. Additional microbial CH₄ contributions may also occur at some mines, consistent with the thermogenic to mixed-thermogenic isotopic signatures observed in available isotope samples (Table 5 and Fig. 9).”

Comment: Is it correct that meteorological data from the Global Forecast System (GFS) is provided on a grid with 28km spacing between points? Was local meteorological data, obtained during the in-situ method flights, used to upscale the GFS data before running the HYSPLIT transport model?

Response: The GFS data used here has a horizontal resolution of approximately 0.25° (~25–30 km). No upscaling or adjustment of the GFS meteorological fields using local in-situ measurements was undertaken.

Instead, HYSPLIT trajectories were used to define the range of potential source regions contributing to the observed curtain. These trajectories were interpreted in conjunction with wind direction measured by the in-situ aircraft, ensuring consistency between large-scale transport patterns and local meteorological conditions.

The emission rates themselves do not incorporate the GFS wind field; it was simply used as a bounding framework to estimate potential upwind sources, particularly those not associated with coal mines.

No modifications have been made to the manuscript.

Comment: Can you further explain the reason for selecting a subset of isotopic analysis results to present in Figure 9? Are the remaining results consistent with a “thermogenic to mixed-thermogenic” source, or does the presented subset contain all valid results? (In the latter case, suggest changing “selected” on page 31 line 761 to “valid” or similar). Aside: the Figure 9 x-axis label, $\delta D-CH_4$ (‰), doesn’t seem to match its description in the caption, δ^2H-CH_4 .

Response: Figure 9 presents $\delta^{13}C$ versus δ^2H (deuterium) for samples where both isotopic measurements were available. The subset shown therefore reflects data availability rather than selective inclusion, and represents all samples with dual isotopic signatures.

Samples for which only $\delta^{13}C$ was measured are not shown in Figure 9; however, these data are consistent with the range observed in the paired $\delta^{13}C$ – δ^2H dataset and do not indicate a different source classification. Due to overlap in $\delta^{13}C$ signatures across source types, these data alone could not be used to further distinguish between sources. We have made this clearer and have corrected the x-axis label to ensure consistency with the caption (δ^2H-CH_4 rather than $\delta D-CH_4$).

Comment: Line 203: Typo in first word – should be “National”.

Response: We have changed accordingly.