Response to referee comments on "The added value of new ground-based observations in improving China's methane emission quantification"

We thank the two referees for their careful reading of the manuscript and the valuable comments. This document is organized as follows: the referees' comments are enclosed by rectangles, our responses are in plain text, and all the revisions in the manuscript and supplementary information are shown in blue. The line numbers in this document refer to the updated manuscript.

Reviewer #2

General comments

I recommend publication following minor revisions. This is a useful paper, and I think the main contribution is clear. The authors are trying to identify where new surface stations would provide the largest increase in sensitivity for methane emissions estimates over China, especially when combined with TROPOMI. The overall framework is interesting, and the results are potentially valuable for network design. My comments are mostly about clarification of the inversion assumptions, interpretation of DOFS, and treatment of errors. The manuscript's core setup is a Bayesian analytical inversion with GEOS-Chem, TROPOMI, and existing ground stations, optimized with simulated annealing. Following are my comments.

Response. We sincerely thank the reviewer for the careful reading of this manuscript and for providing valuable comments. These suggestions have significantly improved the clarity and quality of our work. Below are the answers for each question individually.

1. Please clarify what the DOFS values in Table 1 represent.

In Table 1, I assume the listed DOFS values indicate how much each measurement type or station contributes to constraining emissions over China. If that is correct, please state this more explicitly in the text and/or caption, because as written it is easy to wonder whether these DOFS refer to concentration information, emissions information, or something else. Table 1 currently labels the column as "DOFS in China," and the methods define DOFS as the total number of independent pieces of information provided by the observations, but the connection between those two uses should be stated more plainly.

Response. We sincerely thank the reviewer for pointing out this issue. Specifically, we have made the following changes:

(1) We have renamed the column header to 'DOFS on emissions provided by observations' in Table 1.

(2) We have updated the caption in Line 133. [d] Here, we quantify the individual contribution of each individual station and TROPOMI to constraining methane emissions over China, with each observation source evaluated independently.

(3) We have clarified the text in Line 392. These AK sensitivities and associated DOFS values quantify the observational constraints provided by atmospheric concentration measurements on methane emissions.

2. The treatment of TROPOMI observations needs to be clearer.

In the TROPOMI section and in the discussion of sampling frequency, I was left unsure whether the inversion is

using super observations or whether all valid TROPOMI observations are assimilated individually. This matters for how one interprets the observing system and the assumed error statistics. The paper states that after quality control and spatial regridding, approximately 3.8 × 10^6 valid XCH4 retrievals were obtained over China in 2022, and later discusses large sensitivity to daily, weekly, and monthly sampling frequencies for added surface sites. That raised a basic question for me: are the satellite data being assimilated as individual observations, or aggregated in some way? If the inversion is effectively treating all TROPOMI observations individually, that needs to be stated clearly in the inversion section. If super observations are used, then that also needs to be stated clearly, including how the corresponding observational errors are defined.

Response. Thank you for pointing this out. In this study, our inversion assimilated each TROPOMI retrieval individually. We now make it clear in text.

(1) We have added the following text in Section 2.1 (TROPOMI observations).

Line 83. After quality control and spatial regridding, approximately 3.8 × 10^6 valid **individual** XCH_4 retrievals were obtained over China in 2022 (Fig. 1). **It should be noted that regridding performed here is solely for visualization purposes; in the inversion system, each satellite retrieval is assimilated independently, as in previous studies (Shen et al., 2022; Zhong et al., 2025).**

(2) We have made the following revisions in Section 2.4 (Bayesian analytical inversion).

Line 204. The observation vector ($\boldsymbol{y}$) includes data from TROPOMI and ground stations (see Sections 2.1 and 2.2 for details). Specifically, the inversion assimilates each valid TROPOMI retrieval individually rather than using gridded averages (e.g. super-observations).

3. Please discuss the Gaussian and diagonal error assumptions more carefully.

The inversion framework assumes normal errors and diagonal prior and observational covariance matrices. That is a standard starting point, but for TROPOMI methane retrievals this assumption is not trivial. As Lu Shen knows well, previous GEOS-Chem and TROPOMI inversions often use super observations partly because retrieval and transport errors are not purely independent and normally distributed. I am not asking the authors to redo the inversion, but I do think the manuscript should acknowledge this limitation more directly and explain why the present assumptions are still adequate for the network design exercise.

Response. We thank the reviewer for this insightful comment. We agree that the use of super-observations is also a common practice in atmospheric inversions (Chen et al., 2023; East et al., 2025) to mitigate the influence of spatially correlated observation errors. However, we introduce a regulation factor (γ) to properly quantify the magnitude of observational errors. For TROPOMI satellite observations, we set γ to 0.02. We use γ=1 for ground-based observations, as the sparse network and weekly sampling result in a much lower spatiotemporal density. We have added a new paragraph to Section 2.4 (Bayesian analytical inversion).

Line 180. We assume diagonal structures for the prior ($\mathbf{S_A}$) and observational ($\mathbf{S_O}$) error covariance matrices, a choice that offers significant computational advantages but also entails limitations regarding error characterization. On the one hand, this simplification allows us to evaluate the cost function by processing each individual observation independently, thereby avoiding the construction and inversion of a high-dimensional $\mathbf{S_O}$. On the other hand, we

acknowledge that a diagonal $\mathbf{S_O}$ may not fully account for potential spatial correlations in observation and model transport errors. This assumption can lead to overfitting, particularly when the number of observations far exceeds the number of state vector elements. To address these issues, previous studies have typically employed the parameterization of off-diagonal elements in $\mathbf{S_O}$ to represent spatiotemporal correlations (Liang et al., 2023; Zhang et al., 2022) or the super-observation method to aggregate dense retrievals (Chen et al., 2023; East et al., 2025; Pendergrass et al., 2025). However, these techniques cannot fully eliminate the correlation between observations. To prevent overfitting, a regularization factor (γ) is commonly applied to scale the observational covariance matrix (Brasseur and Jacob, 2017). This factor can be determined by (i) the graph-based L-curve method (Hansen, 2000) or (ii) aligning the posterior state cost function with its theoretical expectation (Lu et al., 2021). Following the latter approach, we use γ = 0.02 for TROPOMI observations. This value is the intermediate one from our tests of 0.01, 0.02 and 0.05, as reported in our previous work (Zhong et al., 2025). In contrast, for sparse ground-based observations, we assume uncorrelated observational errors and adopt a γ value of 1 following previous studies (Lu et al., 2021; Zhang et al., 2022).

4. The discussion of sampling frequency needs more explanation.

The statement that the maximum DOFS enhancement peaks at 2.1, 4.7, and 11.5 for monthly, weekly, and daily sampling caught my attention. Are the authors effectively assuming a daily inversion for those hypothetical new stations? If so, this needs to be spelled out much more clearly in the methods. More broadly, the paper should explain what "daily," "weekly," and "monthly" mean operationally in the inversion, and whether those scenarios assume one representative measurement per day, week, or month, or something more complex. This is important because the resulting gains in DOFS are central to the paper's conclusions.

Response. Thank you for this valuable comment. We have made the following changes to reduce confusion.

(1) The inversion is not performed on a daily basis. Despite we use observations at different sampling frequency, we only solve for the annual mean correction ratios.

Line 172. Note that for all inversions, we only solve for the annual mean correction ratios of the state vector.

(2) We would like to clarify that 'daily', 'weekly', and 'monthly' refer specifically to the assumed observation frequencies of the hypothetical new stations. For the 'daily' scenario, we assume one representative observation per day. Consequently, a station with a daily sampling frequency contributes 365 observations per year to the observation vector (**$\boldsymbol{y}$**), while a weekly one contributes 53 and a monthly one contributes 13.

Section 2.7 at Line 313. We evaluate three sampling frequency scenarios for the hypothetical stations: monthly (observation numbers N=13), weekly (N=53), and daily (N=365). In our experimental setup, we assume one representative measurement is taken at the first hour (00:00 UTC, corresponding to 08:00 local time in China) of each day (or the respective week/month) to maintain temporal consistency across the network. While we utilize daily, weekly, or monthly observation data from these hypothetical stations, the state vector of inversion is optimized at an annual resolution. In general, flask-sampling sites typically operate on a weekly schedule within a globally distributed network (Lan et al., 2021); thus, our analysis primarily focuses on results derived from a weekly sampling frequency.

Section 3.3 at Line 442. Also, the magnitude of these improvements is sensitive to sampling frequency, with maximum DOFS enhancements peaking at 2.1, 4.7, and 11.5 for monthly **(*N*=13)**, weekly **(*N*=53)**, and daily **(*N*=365)** sampling **per station**, respectively.

5. I think the framing around line 255 could be improved.

The paper is identifying optimal locations for new ground-based observation sites by maximizing DOFS, which is fine. But conceptually, adding a measurement will generally increase DOFS under the assumptions of the Bayesian framework used here. So I do not think the central question is whether the added measurement increases DOFS, because by construction it will. The more relevant issue is whether the added site is worth the cost, and whether it adds robust information once potential systematic errors are considered. I think the paper would benefit from a slightly more careful framing here: the optimization is really about maximizing the value of additional sites, not determining whether more measurements help in principle.

Response. We thank the reviewer for this critical insight. We agree that the true value of the study lies in optimizing the location to maximize the observational information. We have added a paragraph at Section 2.7 to explicitly reframe the objective of our optimization.

Line 305. Figure 4 illustrates the workflow of the framework developed in this study to identify optimal locations for a prescribed number of new ground-based observation sites. The approach integrates Bayesian analysis with a simulated-annealing optimization algorithm. **The primary objective of this optimization framework is to maximize the total DOFS (a measure of the number of independent pieces of information provided by the observations) of the integrated observing system across China, which accounts for both existing networks and the proposed new sites.** Since adding any measurement generally increases DOFS by construction within a Bayesian framework, **our optimization focuses on identifying the optimal configurations that provide the maximum information gain.**

6. Please consider discussing information content in terms of posterior error reduction, not just DOFS.

Related to the previous point, I do not think DOFS alone fully captures the value of adding a measurement. What really matters is the reduction in posterior error, especially once systematic errors associated with a new site are considered. A stronger information metric would be the reduction in entropy, or in Rodgers (2000) language, the information gain in bits:

Delta H = 0.5 * log2(|Sa| / |Sx_tilde|) = 0.5 * [log2(|Sa|) - log2(|Sx_tilde|)]

where Sx_tilde is the posterior error covariance including the relevant systematic errors. I do not think the authors need to replace the whole paper with this metric, but I do think some discussion is needed of the distinction between DOFS and actual reduction in posterior uncertainty, especially given the possible role of systematic and correlated errors.

Response. We sincerely thank the reviewer for this insightful suggestion. We agree that DOFS primarily measures the number of independent pieces of information but does not fully account for the uncertainty reduction, especially

when systematic and correlated errors are involved. Following Rogers (2000), we have incorporated a discussion on the information content based on Rodgers (2000). We now use the information gain (ΔH) in bits as a complementary metric to evaluate the measurement value. We have added this analysis in Section 2.5 (Methods) and Section 3.4 (Results) in the revised manuscript.

Section 2.5 at Line 235. However, DOFS do not fully capture the reduction in posterior uncertainty, particularly when systematic or correlated errors are involved. Another useful metric is the information content, which is defined in terms of the Shannon entropy of probability density functions and conceptually linked to thermodynamic entropy (Rodgers, 2000). The information gain of a measurement when the prior error covariance is $\mathbf{S}_A$ and the posterior error covariance is $\hat{\mathbf{S}}$ can be written in bits as:

$$\Delta H = \frac{1}{2}(\log_2|\boldsymbol{S}_A| - \log_2|\hat{\boldsymbol{S}}|) \quad (7)$$

In this study, we primarily rely on DOFS to characterize information content, while ΔH is computed as a supplementary metric to account for posterior error reduction not captured by DOFS alone.

Section 3.4 at Line 543. In addition to DOFS and UR, we further evaluate the Shannon information gain in bits (Fig. S8). In the Bayesian inversion framework, the information gain quantifies the reduction in the entropy of the state vector from the prior to the posterior distribution. TROPOMI alone provides an information gain of approximately 122 bits, increasing to 156 bits when the existing surface network is included. With the addition of hypothetical stations, the information gain increases monotonically from 193 bits (N=5) to nearly 500 bits (N=100). The marginal increase is largest at the initial stage of network expansion and gradually decreases as more stations are added, reflecting diminishing returns as the newly added observations become increasingly redundant with the existing observational constraints.

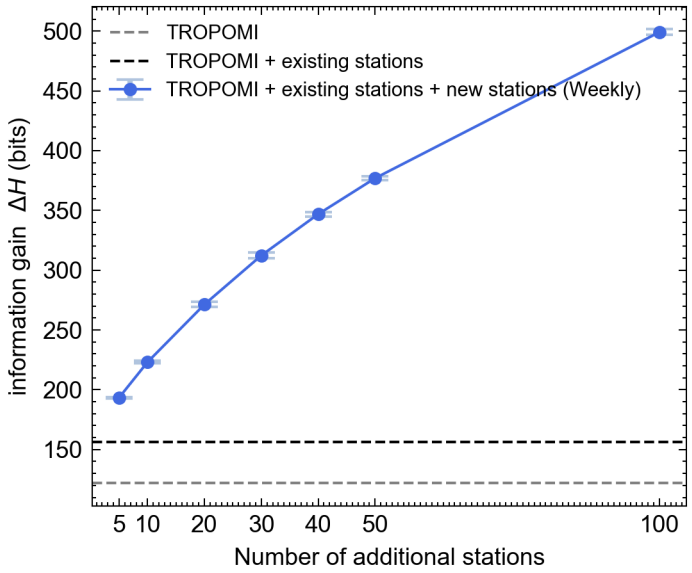

Figure S8. The information gain (ΔH) over China for different observational network configurations. The gray dashed line represents the information gain from TROPOMI observations alone, and the black dashed line represents

TROPOMI combined with the existing surface network. Blue circles show the information gain after further adding hypothetical surface stations (with weekly sampling frequency), with *x*-axis indicating the number of additional stations. Error bars denote the 95% confidence intervals derived from 50 independent simulated annealing experiments.

7. The description of "error" around line 302 is confusing and should be tightened.

I was not always sure whether the paper is referring to observational error, model-observation mismatch, or simply differences between observations and the prior simulation. Those are not the same thing. If the quantity reflects mismatch between observed methane and prior concentrations, then part of that mismatch is signal, not error. On the other hand, if the authors are arguing that unresolved variability at model scale belongs in the observational error term, then that is reasonable, but it needs to be explained more carefully. Right now the wording risks conflating information with error. This matters because the manuscript defines observational error as the aggregate of forward model error, instrument error, and representation error, and then estimates it using a residual-error approach based on model-observation differences.

Response. Thank you for pointing this out this important issue. We use the unresolved variability to quantify the error. We have made the following changes to reduce confusion.

(1) We have clarified the observation error descriptions (Section 2.4 Bayesian analytical inversion).

Line 197. For $\mathbf{S_O}$, we calculate the observational error standard deviations (σ_O) using the residual error method (Heald et al., 2004). The residual error (ε_O) is defined as the difference between observations ($\boldsymbol{y}$) and GEOS-Chem prior simulations ($\boldsymbol{y_a}$), after removing the mean model bias ($\overline{\boldsymbol{y}-\boldsymbol{y_a}}$), as expressed in Eq. (2). The standard deviation of ε_O then gives σ_O (see Table 1).

$$\varepsilon_O = \boldsymbol{y} - \boldsymbol{y_a} - \overline{\boldsymbol{y}-\boldsymbol{y_a}} \quad (2)$$

(2) We have explained the observational errors more carefully in the manuscript.

Line 370. Here, we run GEOS-Chem model at the 0.5°×0.625° resolution, using hourly outputs to estimate observational errors **(sum of instrument error, representative error, and forward model error)** via the residual error method (Eq. 2).

Line 376. **Table 1 summarizes the observation error standard deviations calculated using the residual error method (see Section 2.4) for different observation sources (TROPOMI and surface station observations).** For ground-based observations representing boundary-layer concentrations (ObsPack and CMA), **observation error standard deviations** range from 20 to 82 ppb and generally increase from background to urban sites near emission sources. Larger observation errors at urban sites reflect the greater variability in airflow (alternating between polluted and background air) and the difficulty of capturing high-frequency transport features at the model's current resolution. For column-integrated measurements (TCCON and TROPOMI), **observation error standard deviations** are typically 10-15 ppb. This smaller uncertainty arises because most column methane originates from the free troposphere, where variability is lower, and boundary-layer fluctuations contribute only a small fraction of the total column signal. Based on Table 1, we use the mean **observation error standard deviation** from six urban sites (TAP,

AMY, SDZ, LFS, JSA, and LAN) as a representative estimate for potential new stations (σ_O = 65 ppb). This is because new ground sites are likely to be deployed in eastern China (see Fig. 3d), where emissions are high, so the urban error statistics provide a realistic proxy for expected observational uncertainty in these regions.

8. The relation between large "errors" and large DOFS deserves explanation.

At line 314 and in Table 1, I may be reading this incorrectly, but it looks like some sites with relatively large assumed errors also have relatively large DOFS. If that is right, it would help to explain this explicitly. My guess is that proximity to strong source regions and stronger Jacobian sensitivity can outweigh larger observational error, but the reader should not have to infer that. A short explanation would help. The paper already hints at this by noting that stations near emission sources provide greater constraints, so this should be easy to clarify.

Response. We thank the reviewer for this insightful comment. The reviewer is correct that proximity to strong source regions leads to a larger Jacobian sensitivity, which can outweigh larger observational errors and result in higher DOFS. However, we would like to clarify that the DOFS is also heavily influenced by the actual number of valid observations available at each station. As shown in Table 1, sites with the highest DOFS among all CMA sites (e.g. SDZ and LAN) are characterized by significantly larger datasets of actual measurements.

We have added a new Figure S4, and made the following changes in text.

Section 3.2 at Line 395. Among all CMA sites, the highest sensitivities occur at SDZ (northern China) and LAN (Yangtze River Delta), both located in high-emission regions. Despite their relatively large observation error standard deviations (80 ppb and 82 ppb), these sites still yield high DOFS values. Theoretically, since the prior error ($\mathbf{S_A}$) and the regularization factor (γ) are kept consistent across all sites, variations in DOFS are primarily driven by the Jacobian matrix ($\mathbf{K}$) and the observation error ($\mathbf{S_O}$). For these two sites, the higher DOFS can be attributed to their stronger local emission fluxes (i.e., higher concentration sensitivity, see Fig. S4) and larger observation numbers, which allow for a greater accumulation of information in the term of the $\gamma\mathbf{K}^T\mathbf{S_O}^{-1}\mathbf{K}$ term. Conversely, site such as YON lacks prior emissions within their respective grid cell, but maintains high DOFS values. This is because it can capture transported signals from adjacent sources; furthermore, its high-frequency observations provide robust constraints.

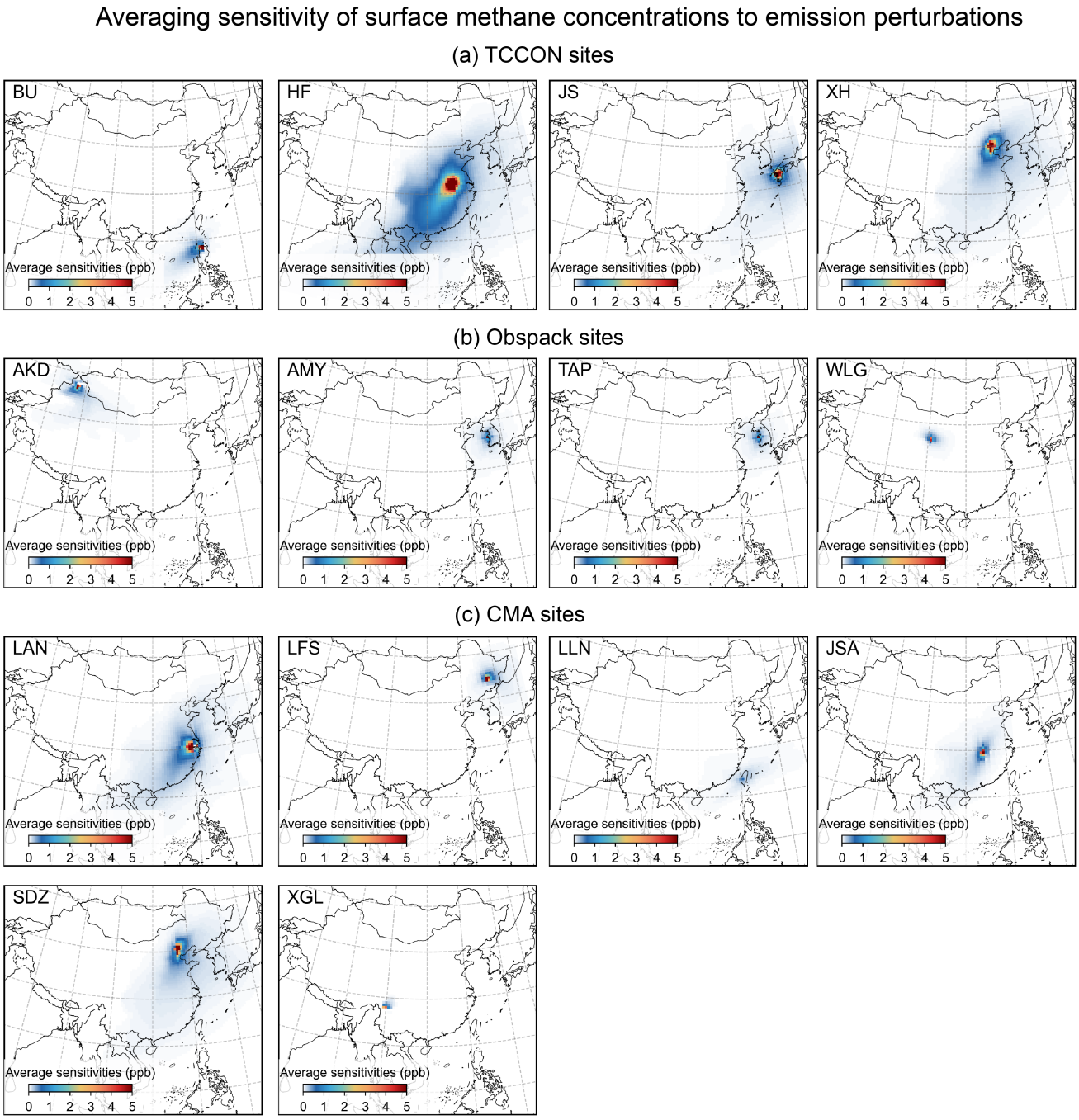

Figure S4. Hourly averaged sensitivities (Jacobians) of surface methane concentrations to localized emission perturbations. Sensitivities are derived from GEOS-Chem simulations by perturbing emissions and computing the concentration changes. Note that for maritime sites (e.g., YON, DSI and TPI), the absence of prior emissions within their respective state vector precludes direct local perturbation. However, these sites provide essential constraints on upwind emissions from adjacent terrestrial regions via regional transport.

9. Equation 7 should include an important caveat about source attribution.

For the sectoral attribution framework following Hancock et al. ACP 2025, I strongly recommend adding a sentence noting that this method assumes the relative contributions of each sector to the total emissions in a given grid cell are

correct, which introduces an additional source of uncertainty in the sectoral attribution of inversion results. Although the high resolution of the inversion reduces the impact of this assumption relative to coarser approaches, the posterior attribution to sectors still depends on the spatial allocation in the prior inventories. I also recommend citing Cusworth et al. (2021), which presents a Bayesian framework for deriving sector-based methane emissions from top-down fluxes and explicitly discusses uncertainty structure in sector attribution.

Suggested citation text for this section:
"Following Hancock et al. (2025), this method assumes that the relative contributions of each sector to the total emissions in a given grid cell are correct, which introduces an additional source of uncertainty in the sectoral attribution of inversion results. Although the high resolution of our inversion reduces the impact of this assumption compared with coarser-resolution approaches, our ability to attribute posterior emissions to individual sectors remains dependent on the spatial allocation of emissions in the prior inventories."

Relevant citation links:
Hancock et al. (2025): https://acp.copernicus.org/articles/25/797/2025/
Cusworth et al. (2021): https://www.nature.com/articles/s43247-021-00312-6

Response. Thanks for your helpful suggestion. We agree that sectoral attribution based on prior relative contributions introduces additional uncertainty. We have added the suggested discussion and cited these two papers to explicitly acknowledge the dependency of posterior attribution on the prior spatial allocation.

Line 260. Because our inversion directly constrains only total methane emissions rather than individual sectors, this method attributes sectoral methane emissions based on their prior fractional contributions within each grid cell. This assumes that this prior sectoral contribution is correct in a given grid cell (Hancock et al., 2025). While the high spatial resolution of our inversion helps mitigate the impact of this assumption compared to coarser resolutions, this approach nevertheless introduces an additional source of uncertainty in the sectoral attribution of the posterior results (Cusworth et al., 2021).

10. A couple of assumptions deserve a bit more discussion in the limitations.
The paper assumes diagonal observational and prior covariance matrices, and it also assigns a fixed observational error to hypothetical new stations based on the mean error from six urban sites. Both are reasonable practical choices, but both could affect the resulting optimization. I suggest briefly discussing how sensitive the site selection may be to these assumptions. Related to this, the CMA recent data are not publicly available, so the 2022 analysis applies error characteristics from earlier years. That is understandable, but it is still a limitation worth stating explicitly. The manuscript does mention some limitations in the Discussion, because they directly affect the information metrics that drive the optimization.

Response. Thanks for pointing it out. We have implemented three specific modifications in the revised manuscript. In particular, the following descriptions have been incorporated:

(1) We have tested the influence of different choices of gamma and observational error on the observational

constraints from individual new ground sites (Section 3.3). Furthermore, to investigate the sensitivity of the observational constraints to the regularization factor and the observation error standard deviation when hypothetically adding a single surface station, we perform additional sensitivity experiments with γ = 0.7 (referencing Zhang et al., 2022) and observation error standard deviations of 75 ppb and 55 ppb (corresponding to 65±10 ppb). As shown in Fig. S5, the spatial distribution of enhancement hotspots remains largely unchanged compared to Fig. 6. Larger σ_O or a smaller γ results in weaker observational constraints.

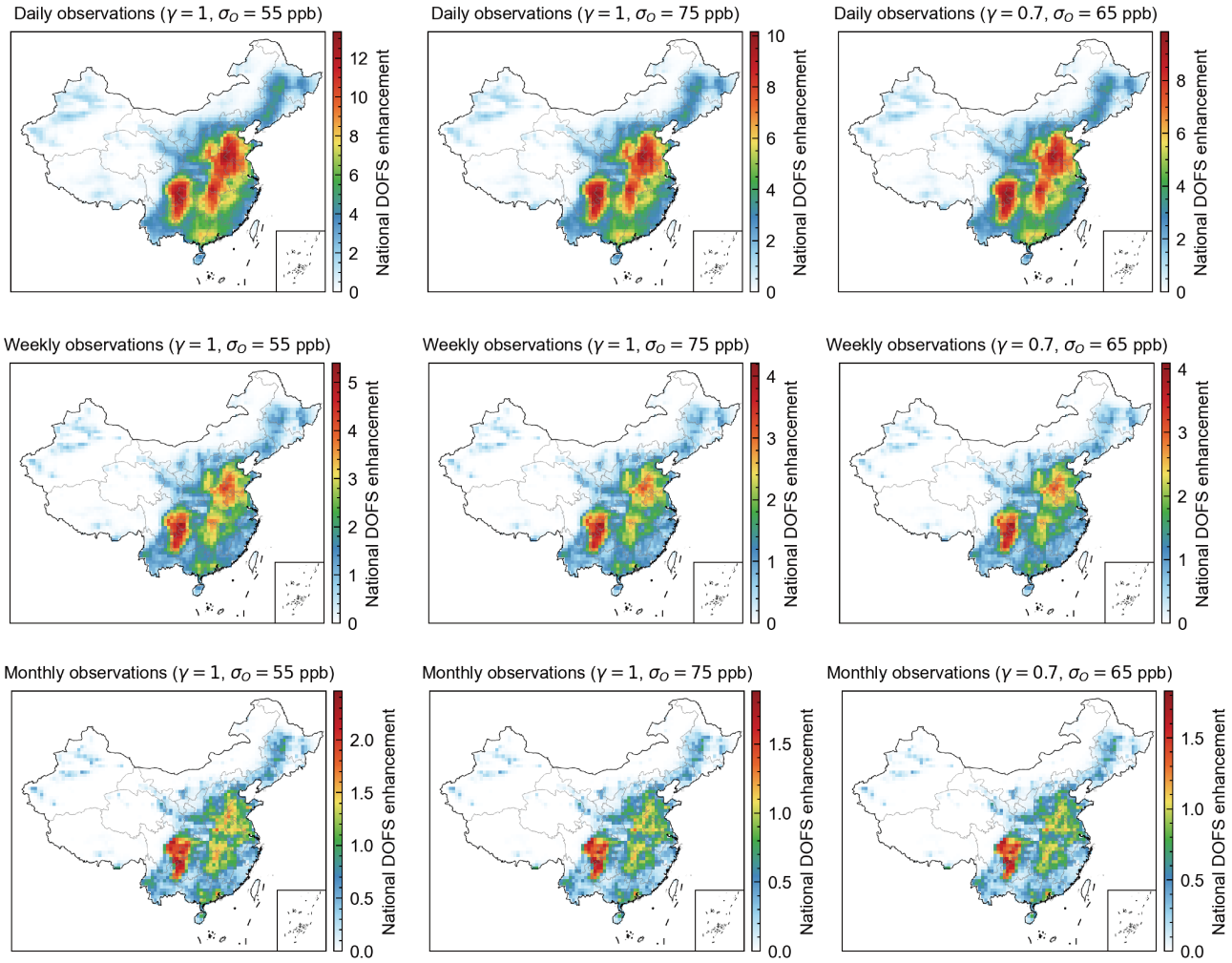

Figure S5. The degrees of freedom for signal (DOFS) enhancement in China by adding one ground station with different temporal frequencies and parameter selection in the inversion. The results are evaluated under different configurations of the regularization factor (γ = 0.7 and 1) and observation error standard deviations (σ_O= 55, 65 and 75 ppb).

(2) We have clarified the observation error sections and assumed a variable observation error for new hypothetical sites using the RSSD method, performing 10 independent simulations to examine the final results.

Line 196. For $\mathbf{S_O}$, we calculate the observational error standard deviations (σ_O) using the residual error method (Heald et al., 2004). The residual error (ε_O) is defined as the difference between observations ($\boldsymbol{y}$) and GEOS-Chem prior simulations ($\boldsymbol{y_a}$), after removing the mean model bias ($\overline{\boldsymbol{y}-\boldsymbol{y_a}}$), as expressed in Eq. (2). The standard deviation

of ε_O then gives σ_O (see Table 1).

$$\varepsilon_O = \boldsymbol{y} - \boldsymbol{y_a} - \overline{\boldsymbol{y} - \boldsymbol{y_a}} \quad (2)$$

Line 210. In part of the analysis, to account for the potential dependence of observational error on methane concentrations, we calculate the relative residual standard deviation (RRSD) as the standard deviation of the residual error (σ_O) divided by the modelled annual mean methane concentration in each grid cell. We adopt a reference RRSD of 3.3%, derived from six existing urban sites (Fig. S2). Thus, the hourly observational error ($\varepsilon_{O,i}$) is calculated as:

$$\varepsilon_{O,i} = RRSD \times y_{a,i} \quad (3)$$

where $y_{a,i}$ is the modeled methane concentration at hour i.

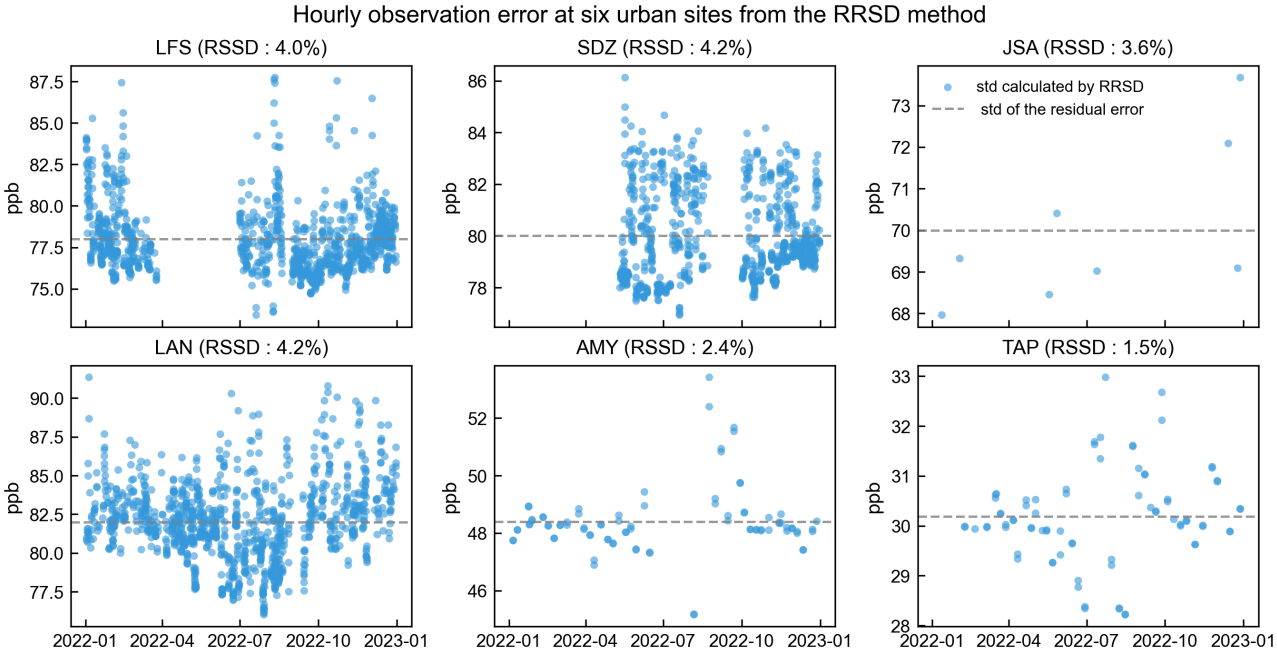

Figure S2. Hourly error at six existing urban stations derived from the relative residual standard deviation (RRSD) method. Gray dashed lines represent the standard deviation of the residual error. Blue dots denote the observation-specific error standard deviations calculated using the RRSD method (see Section 2.4 for details).

Line 479. We also solve for the optimal locations using the error standard deviations calculated from the RRSD method under a weekly sampling assumption (Fig. S6). The resulting spatial distribution of the newly added stations is largely consistent with that obtained using a fixed standard deviation of 65 ppb.

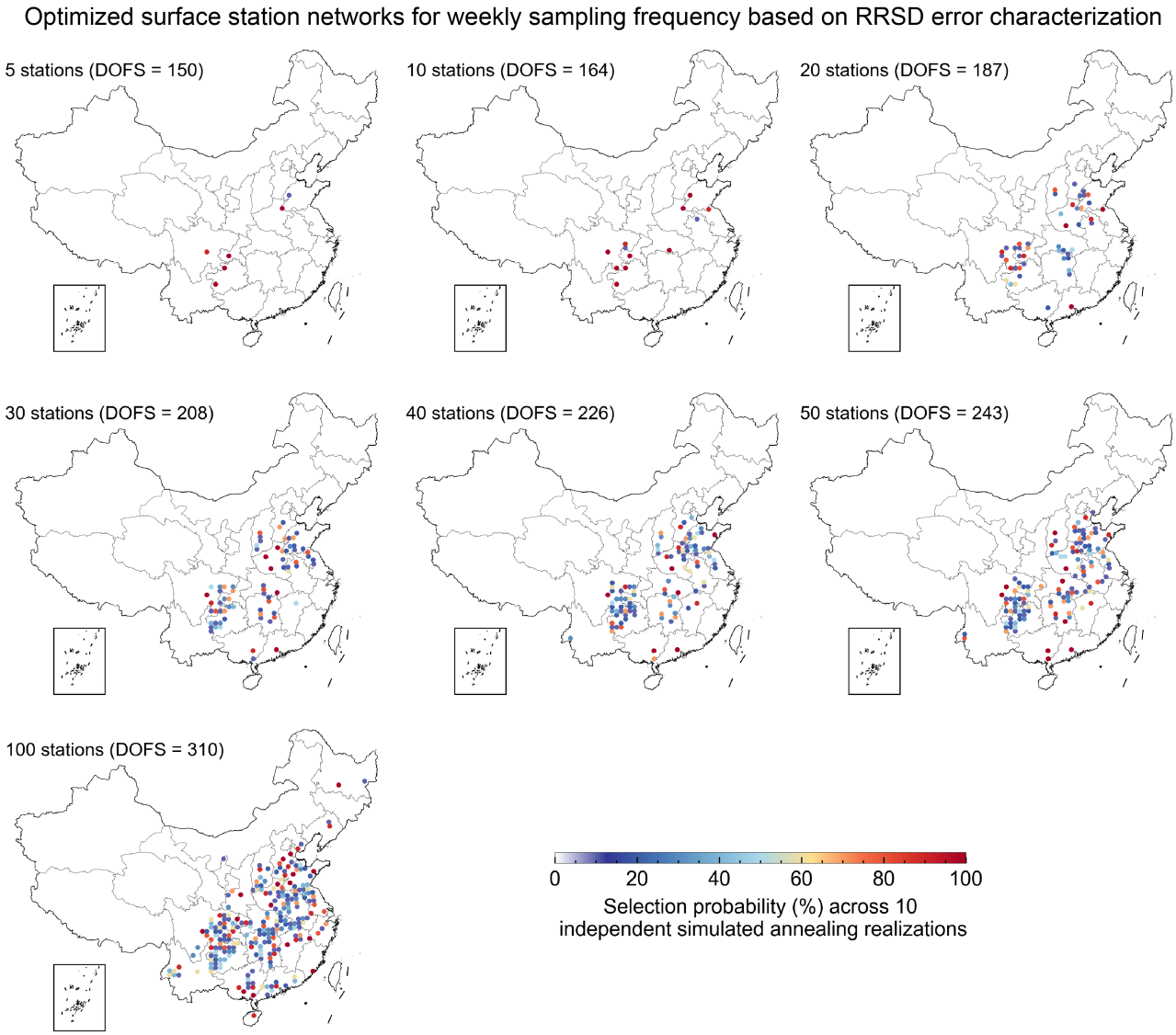

Figure S6. Optimized network expansion under weekly sampling when observation error is calculated by the relative residual standard deviation (RRSD) method. Panels show the selection probability (%) of each grid cell for a new site across 10 independent simulated annealing realizations, for network expansions of 5 to 100 new sites. The corresponding increase in the degrees of freedom for signal (DOFS) is displayed within each panel.

(3) We added a discussion on the limitations of applying error characteristics from earlier years for the CMA data (Section 4). While our network optimization framework represents a significant advance in monitoring design, this study is subject to several limitations. **These uncertainties primarily stem from inversion assumptions and the limited availability of observational data. First, our framework assumes diagonal observational and prior covariance matrices. This assumption of independent errors is likely unrealistic and can lead to overfitting. We apply a regularization parameter (γ) to avoid overfitting. For ground-based stations, we test two values (0.7 and 1) and obtain consistent results (Section 3.3). Second, due to the lack of publicly available CMA recent data, the 2022 analysis relies on error characteristics from earlier years. Since the calculation of DOFS requires only error information rather than actual observations, this approach assumes that the error characteristics constant over time, which may introduce additional uncertainty.** Third, the spatial patterns of prior inventories

can substantially affect the estimates of DOFS (Chen et al., 2022). This may bias site selection toward regions where prior emissions are potentially overestimated. Furthermore, our current framework does not consider complementary information from other satellite datasets. In particular, thermal infrared (TIR) sensors such as IASI offer distinct vertical sensitivity profiles (Siddans et al., 2017), and China's Gaofen-5B also has the potential to enhance detection capabilities (He et al., 2024). Incorporating these observation platforms would represent meaningful directions for future improvements to the network design strategy.

11. The uncertainty-reduction metric could be brought more centrally into the discussion.
The manuscript includes relative uncertainty reduction in addition to DOFS, and I think that helps connect the optimization to what readers care about physically. Since DOFS is not the same thing as total information gain or practical reduction in posterior uncertainty, I would encourage the authors to emphasize the UR results a bit more in the interpretation. In my view, this would strengthen the paper. The manuscript already computes UR, maps it, and notes that local uncertainty reductions near newly added stations can exceed 90%, so this material is already there and just needs slightly more emphasis.

Response. We appreciate the reviewer's suggestion to emphasize the uncertainty reduction (UR) results. We have re-evaluated our calculations and identified a computational oversight in the previous version, where the UR values were incorrectly processed. We have now corrected these calculations using the standard deviation from prior to posterior. Accordingly, we have updated the corresponding figures and text throughout the manuscript. We have also added a regional mean UR analysis to provide a more detailed evaluation of the network's performance.

Line 529. Figure 9 illustrates the relative uncertainty reduction (ΔUR) achieved by a representative case (weekly sampling frequency) selected from 50 independent simulated annealing experiments (as presented in Fig. 7), quantified by comparing posterior uncertainties before and after ground-based network expansion (Eq. 12). The locations of the newly added stations are indicated by the black circles in the figure. **As shown in Fig. 9, local relative uncertainty reductions near newly added stations can exceed 50-70%, reflecting the ability of ground-based measurements to directly sample methane enhancements within the boundary layer and to better constrain nearby sources. Such an impact decreases with distance, with uncertainty reductions of 10-50% within 100-200 km and of < 10% beyond ~200 km. To evaluate the posterior uncertainty reduction at a regional scale, we calculate the regional mean uncertainty reduction (Fig. S7) across the four regions defined in Fig. 1. The posterior uncertainty reduction exhibits significant spatial heterogeneity across regions as the number of added sites increases from 5 to 100. The Southeast and Southwest regions show the most substantial and continuous improvements, with reductions increasing from 9.2% to 21.3% and 4.3% to 13.5%, respectively. Conversely, the added value in northern regions is much smaller, reflecting the already strong observational constraints provided by the dense coverage of existing TROPOMI data.**

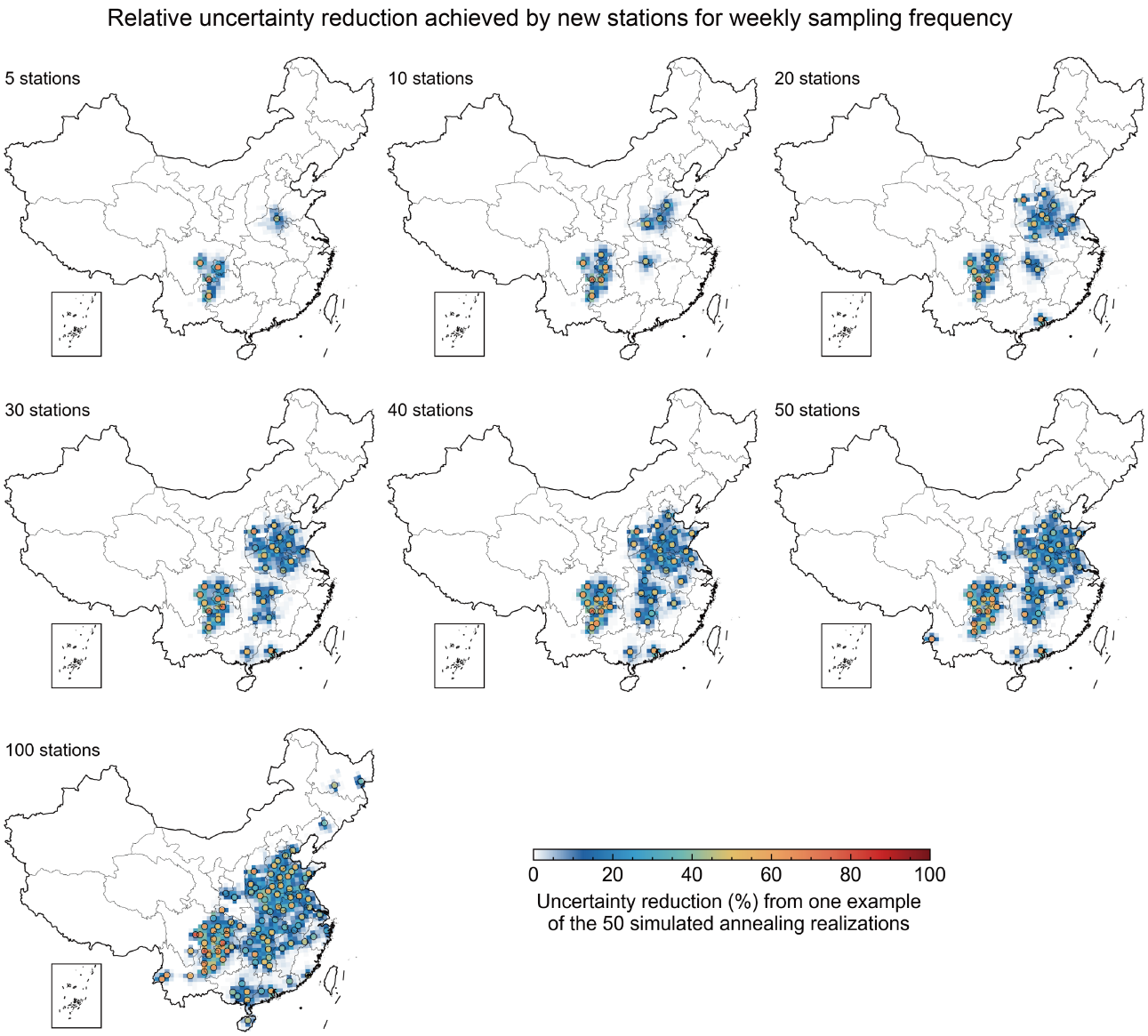

Figure 9. Spatial distribution of the relative uncertainty reduction achieved by expanding the existing observational network under weekly sampling frequency. The figure shows the result from one case of the 50 simulated annealing realizations for different numbers of new stations (ranging from 5 to 100). The existing observational network consists of TROPOMI satellite observations combined with the established ground-based stations. The relative uncertainty reduction (ΔUR) is calculated per 0.5°×0.625° grid cell. Newly added station locations in each scenario are marked with black circles.

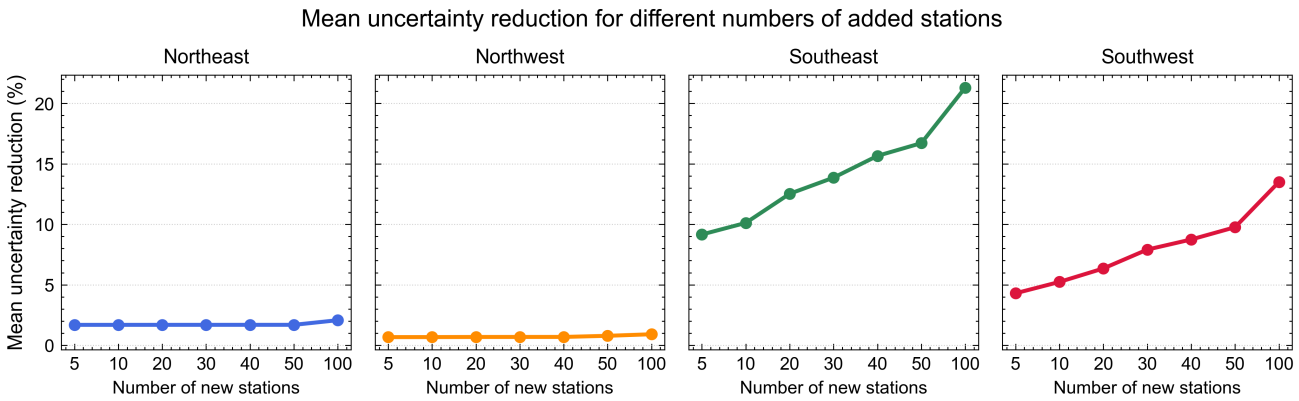

Figure S7. Regional mean uncertainty reduction contributed by the existing observational network and the addition of new surface sites. The hypothetical new stations (ranging from 5 to 100) are assumed to have a fixed weekly sampling frequency.

12. Please clarify exactly what level of emissions is being optimized and interpreted.
At different points the paper discusses national DOFS, regional DOFS, and sectoral averaging kernel sensitivities. Those are all useful, but the transitions between them are a bit quick. A short clarifying paragraph would help readers keep straight whether the optimization is targeting grid scale emissions, national totals, or sectoral attribution, and how those relate. Right now the manuscript moves from a state vector of native grid cells, clusters, and boundary conditions, to national DOFS, then regional DOFS, then sectoral AK sensitivities. I think the science is fine, but the reader would benefit from a clearer roadmap.

Response. Thanks for your helpful suggestion and reminder. We have added Section 2.8 (Evaluation of the performance of the new network) to clarify what level of emissions is being optimized and interpreted.

Line 345. In summary, our inversion system optimizes a state vector (including native grid cells, clusters and boundary conditions) for the annual mean correction ratios. To evaluate the performance of the newly designed ground-based networks, we employ a multi-scale assessment framework (including DOFS, UR, ΔH and $\mathbf{A}_{red}$) that spans from individual grid cells to national-level (native grid cells and clusters within China) contributions. First, we define national DOFS as the primary metric and objective function for our simulated annealing algorithm to identify the optimal ground-based network (see Section 2.5 and 2.6). Second, we evaluate the enhancement in national DOFS resulting from the addition of a single new site at different locations across China (see Section 3.3). Third, based on the optimized networks from the simulated annealing experiments (see Section 3.4), we calculate: (i) regional DOFS by aggregating the diagonal elements of the averaging kernel matrix ($\mathbf{A}$) over specific domains (see Section 2.5); (ii) grid-scale and regional-scale uncertainty reduction (UR) based on prior and posterior error standard deviations (see Section 2.8); and (iii) information gain at the national scale to better capture the overall reduction in uncertainty (see Section 2.5). Finally, we derive sector-specific averaging kernel sensitivities ($\mathbf{A}_{red}$) by employing a summation matrix $\mathbf{W}$ to project the high-dimensional $\mathbf{A}$ matrix onto sector-specific subspaces and grid-scale (see Section 2.5, 3.5).

Minor point

The statement that areas south of 33° N account for only 3.8 × 10^5 valid retrievals, about 11% of those in northern China, is useful and helps motivate the paper. I would keep it, but I think it would be even stronger if the authors tied this more directly to the network-design result, namely that the optimized new stations repeatedly cluster in southwestern and southern China because that is exactly where the observing gap is most consequential.

Response. Thank you for pointing this out. We have added the following description:

Line 91 (Section 2.1). This pronounced observational gap in southern China suggests that strategic placement of new ground-based stations in this region may be essential to compensate for the limited satellite constraints.

Line 472 (Section 3.4). This spatial clustering directly addresses the critical observing gap previously identified, where the number of valid TROPOMI retrievals in southern China (south of 33°N) is only 11% of that in northern China.