

Author’s Response to Reviewer #2: GHGPSE-Net: A method towards spaceborne automated extraction of greenhouse-gas point sources using point-object-detection deep neural network

Yiguo Pang^{1,2}, Denghui Hu¹, Longfei Tian¹, Shuang Gao¹, and Guohua Liu^{1,2}

¹Innovation Academy for Microsatellites of Chinese Academy of Sciences, Shanghai, China

²University of Chinese Academy of Sciences, Beijing, China

Contact author: Yiguo Pang (pangyiguo21@mailsucas.ac.cn)

Corresponding author: Guohua Liu (liugh@microsate.com)

October 20, 2025

We sincerely thank the editor and reviewers for the valuable comments and suggestions. Our detailed point-by-point responses to reviewer #2 are given as follows, where the reviewer’s comments are in *blue italics* and our responses are in normal text.

Point-by-point Response

This work applies deep learning to satellite snapshots of CO₂, performing point source detection, localisation and quantification. As the authors identify, most work in this area focuses on plume identification or source location, without performing quantification directly.

We appreciate your recognition of our work. Indeed, most existing studies focus on plume detection and the source localization are often performed manually based on the detected plumes. The quantification process and plume detection are often treated as separate processes. This issue becomes significant when considering large volume of moderate satellite observations, where each pixel covers a broad area. Such data requires both precise localization and efficient processing.

The authors present an innovative solution, where a CNN is trained to identify sources as hotspots, which then get detected and quantified using a statistical Gaussian Kernel fitting method.

We thank the reviewer for the summary of our work. You pointed out the key components of our approach, including the CNN-based hotspot identification and the Gaussian Kernel fitting for quantification. Though, it is worth noting that the Gaussian kernel fitting is based on least

squares fitting rather than statistical fitting. This offers better precision than traditional Non-Maximum Suppression (NMS) method, as demonstrate in Fig. 6. We estimate the source parameters by minimizing the residuals between the UNet-predicted heatmap and the Gaussian kernel model.

The choice of Gaussian kernel is rather empirical and is selected for several reasons. (1) To represent point objects in a CNN-generated heatmap, a smooth 2D function is necessary. Directly using delta functions, such as one-hot binary masks, results in sparse gradients and poor convergence. (2) The Gaussian kernel function has a simple mathematical form and is used by previous studies to represent object size and location [Law and Deng, 2019]. (3) The form of Gaussian kernel is aligned with the point source parameters, where the center and peak value have physical meanings.

The authors also introduce a greenhouse-specific augmentation method, varying the linear coefficients for each source, that presents a valuable contribution to the area of GHG+deep learning.

We appreciate the reviewer’s commnets. It is worth noting that the augmentation method is directly inspired by that used in Dumont Le Brazidec et al. [2024]. Here, we generalize the problem in a linear combination and write it into a clear mathematical form. Furthermore, we also use statistical analysis to decide the linear coefficient ranges, instead of empirically defined. We suppose this form of GHG-specific data augmentation method can be intuitive and useful for future studies. For example, as we mentioned in the manuscript, the noise term can be more complex than naive Gaussian noise. Future studies may consider adding instrument-specific/scene-specific noise patterns into the augmentation process.

Clarifications required:

The datasets used should be summarised more cohesively, as at the moment it is hard to follow 1) what regions, areas and time periods are covered for each dataset 2) the number of samples in each dataset 3) whether the test datasets were augmented. The number of samples added in the augmentation needs to be specified.

Thank you for your constructive suggestions. We agree that a clear summary of the dataset generation process is important for better clarification, as the entire generation process is rather complicated. We will improve the description of the datasets in the revised manuscript. Temporally, we would answer your questions as follows:

(1) The XCO_2 data before augmentation is generated by WRF-GHG simulations. We generated a XCO_2 snapshot for each tracer channel every hour. Based on the synthetic XCO_2 snapshots, we further generated instrument-specific datasets (HGET, HGET-3ppm, UCPI, etc.). So, in a nutshell, the regions and time periods covered by each dataset are the same as those of the WRF-GHG simulations, which are described in detail in Section 2.1.1. Further details, including source codes, configuration files, initial conditions, boundary conditions and emissions files, are provided in a Zenodo repository (<https://zenodo.org/records/17337441>).

(2) The number of XCO_2 snapshots before augmentation is 2,400. For each instrument scenario, we generated 24,000 samples using the method described in Section 2.1.2, with data augmentation (Section 2.1.3) from the original snapshots. Each dataset is divided into training, validation, and testing sets in a 3:1:1 ratio (Section 2.2.3).

(3) The data augmentation is performed before the construction of datasets, due to efficiency consideration and simplicity. Therefore, the test datasets are also augmented. Actually, all the datasets used in this study are generated with data augmentation, except for the control group’s training stage at the beginning of Section "Generalization evaluation using SMART-

CARB dataset and OCO-3 observation".

The architecture of the model needs to be outlined more clearly, including how it was tuned, the learning rate, loss functions etc.

Thank you for your comments. The general architecture of the model is described in Section 2.2.1 and the training setups are described in Section 2.2.3. We will further improve the description of the model training process. Specifically, we used the Adam optimizer, with an initial learning rate of 0.001, which is a classical choice in deep learning tasks. The loss function is the Mean Squared Error (MSE) between the predicted heatmap and the ground truth heatmap, as described in Section 2.2.3. Further details of the model can be referred to a Zenodo repository (<https://zenodo.org/records/16751293>).

It is not clear to me if the resolution of the model was kept constant across the different instruments, or adapted to each instrument.

Thank you for your question. As we described in Section 2.1.2, the resolution of the XCO_2 data follows the instrument pixel size and image size. We currently focus on building a single-instrument-specific model, instead of a general model that can adapt to different instruments. For each instrument scenario, the models are trained and evaluated using different datasets. We suppose a further clarification of the dataset generation process will help to understand this point. We will improve the description in the revised manuscript.

Is it possible that the significantly lower skill on the SMARTCARB dataset is due to a different transport model?

That's an insightful point. The SMARTCARB dataset is generated using the COSMO-GHG model (Kuhlmann et al. [2019]), which is different from the WRF-GHG model used in our training dataset. The different transport models may lead to different plume shapes and distributions as discussed by Brunner et al. [2023]. We treat the emission sources as ground sources, and the plumes are mostly confined near the ground. This is slightly different from the elevated sources used in the SMARTCARB dataset [Brunner et al., 2019, Kuhlmann et al., 2019]. Our model may underestimate the plumes entering the higher altitudes, as we discussed in the manuscript. We also found a false positive case where a plume structure that may have escaped the boundary layer was identified as an independent plume (shown in Fig. S15). Similarly, other biases among different transport models may also contribute to the performance degradation. However, further analysis is difficult and seems beyond the scope of this study. Further investigations may be needed to fully understand these biases, and they are currently less discussed in the GHG remote sensing community. We will extend the discussion in the revised manuscript.

Would other inputs to the CNN, like the wind direction or the location of known sources, improve predictive skill?

Thank you for your suggestions. This point is also raised by Reviewer #1, and thus we share the same response here. Currently, GHGPSE-Net does not require the 2D wind field as input and performs source extraction solely using XCO_2 observations. Indeed, additional input to the model, such as wind field, NO_2 concentration, will enhance the model performance, as demonstrated by Kuhlmann et al. [2020] and Dumont Le Brazidec et al. [2024]. Additional input, such as cloud cover, surface type and land-ocean mask, may be helpful to perform quality control. However, the inclusion of additional input will increase the model complexity and make the analysis more difficult. In this initial study, we focus on developing a basic framework for GHG point source extraction using only XCO_2 observations. Future work will explore the

inclusion of additional inputs and improve its real-world applications. We will further clarify this point in the "Discussions and Conclusion" section.

The Gaussian kernel fitting is demonstrated to improve when the kernel size is the same as the instrument pixel size, for the one instrument tested. Do you expect performance to improve for other instruments when the same process is applied? It is not clear what kernel size you applied for evaluation for each instrument.

That's an accurate understanding. The kernel size achieves optimal performance in a moderate size. As the kernel that is too wide tends to blur local features, making the kernel fitting less accurate. In contrast, a kernel that is too narrow tends to bring in overly localized gradients, hindering parameter updates in the training. The initial experiment using HGET shows that the optimal kernel size is close to the instrument pixel size. We suppose this conclusion can be generalized to other instruments, as similar forms and reasons also apply to other instruments. Moreover, it is worth noting that the optimal kernel size may not exactly match the instrument pixel size, as other factors such as instrument noise, scene complexity may also affect the optimal kernel size. The complexity of the deep learning model further makes theoretical analysis difficult. However, as the core goal of this work is to demonstrate the feasibility of using a point-detection neural network for GHG point source extraction, further large-scale controlled experiments for the choice of kernel size can be rather petty and make the literature redundant.

Regarding the kernel size (σ value) used for other instruments, we used a kernel size of 0.5 km for the baseline HGET scenario and HGET-3ppm scenario. For the UCPI and UCPI-3ppm scenarios, we used a kernel size of 2 km. For the OCO-3 scenario, we used a kernel size of 2.25 km. A controlled experiment using varying kernel sizes is only applied to the HGET scenario.

Typos and minor corrections:

Figure 5 shows, from the text, two days with large absolute errors. Clarify in the caption/figure labelling that these two days are not representative of the usual error distribution.

Thank you for your constructive suggestion. We will add it to the caption in the next revision. It is worth noting that these errors are mostly observed downwind of the city center of Shanghai and heavy industrial areas near the Yangtze River. We've attached a spatial map of the errors in Fig. S13, which may help to understand the error distribution. We will further clarify this point in the revision.

Line 39 ("and The new generation...")

Thank you for pointing out this typo. We will correct it in the revised manuscript.

Line 151: areas (not area)

Thank you for pointing out this typo. We will correct it in the revised manuscript.

L342

Thank you for pointing out this typo. "The" should be "the", and we will correct it.

Supplement S3.1, title: Definition (not defination)

Thank you for pointing out this typo. We will correct it in the revised manuscript.

References

- Dominik Brunner, Gerrit Kuhlmann, Julia Marshall, Valentin Clément, Oliver Fuhrer, Grégoire Broquet, Armin Löscher, and Yajka Meijer. Accounting for the vertical distribution of emissions in atmospheric CO₂ simulations. *Atmospheric Chemistry and Physics*, 19(7): 4541–4559, April 2019. ISSN 1680-7316. doi: 10.5194/acp-19-4541-2019.
- Dominik Brunner, Gerrit Kuhlmann, Stephan Henne, Erik Koene, Bastian Kern, Sebastian Wolff, Christiane Voigt, Patrick Jöckel, Christoph Kiemle, Anke Roiger, Alina Fiehn, Sven Krautwurst, Konstantin Gerilowski, Heinrich Bovensmann, Jakob Borchardt, Michal Galkowski, Christoph Gerbig, Julia Marshall, Andrzej Klonecki, Pascal Prunet, Robert Hanfland, Margit Pattantyús-Ábrahám, Andrzej Wyszogrodzki, and Andreas Fix. Evaluation of simulated CO₂ power plant plumes from six high-resolution atmospheric transport models. *Atmospheric Chemistry and Physics*, 23(4):2699–2728, February 2023. ISSN 1680-7316. doi: 10.5194/acp-23-2699-2023.
- Joffrey Dumont Le Brazidec, Pierre Vanderbecken, Alban Farchi, Grégoire Broquet, Gerrit Kuhlmann, and Marc Bocquet. Deep learning applied to CO₂ power plant emissions quantification using simulated satellite images. *Geoscientific Model Development*, 17(5):1995–2014, March 2024. ISSN 1991-959X. doi: 10.5194/gmd-17-1995-2024.
- Gerrit Kuhlmann, Valentin Clément, Julia Marshall, Oliver Fuhrer, Grégoire Broquet, Christina Schnadt-Poberaj, Armin Löscher, Yajka Meijer, and Dominik Brunner. SMARTCARB – Use of satellite measurements of auxiliary reactive trace gases for fossil fuel carbon dioxide emission estimation. Technical report, Zenodo, January 2019.
- Gerrit Kuhlmann, Dominik Brunner, Gregoire Broquet, and Yajka Meijer. Quantifying CO₂ emissions of a city with the Copernicus Anthropogenic CO₂ Monitoring satellite mission. *Atmospheric Measurement Techniques*, 13(12):6733–6754, December 2020. ISSN 1867-1381. doi: 10.5194/amt-13-6733-2020.
- Hei Law and Jia Deng. CornerNet: Detecting Objects as Paired Keypoints, March 2019.