Dear Referee,

We thank the referee for taking the time to provide such a constructive and thorough review of our manuscript (GMD-2023-173). According to the suggestive comments, we have made some modifications to the manuscript, and the responses are listed below. To guide the review process, comments from the referee and original texts in the manuscript are presented in black, our responses are in blue, and any text modifications made to the manuscript are highlighted in red italics. Links are provided below for easy navigation in the document.

We are looking forward to your reply.

Best regards,
Yours sincerely
Sheng Fang

General comments

This study proposes a novel approach to the source reconstruction of atmospheric radionuclide emissions in non-stationary emission scenarios. By moving away from the unrealistic assumption of constant emissions and developing a method for spatiotemporally decoupled source reconstruction, it effectively leverages the fact that variations in emission rates significantly impact observations. The methodology involves training machine learning models with the XGBoost algorithm and determining detailed temporal variations in emission rates using the PAMILT algorithm. The paper makes a significant contribution to the field of atmospheric radionuclide emission source reconstruction. The proposed methodology offers an effective means for accurately localizing sources and estimating emission rates in non-stationary scenarios, presenting a promising framework for future research and applications.

[1] The utilization of a temporal sliding-window average filter is commendable. However, elucidating the criteria for feature selection and the impact of varying combinations of observation sites on source estimation would enhance the paper.

[2] Validating the proposed method against the SCK-CEN ^{41}Ar field experiment data underscores its efficacy and applicability. Nonetheless, conducting further validation studies under diverse scenarios and conditions would enrich our understanding of the method's applicability and limitations. It is recommended to include additional case studies involving different types of releases and weather conditions to assess the method's efficiency and adaptability more comprehensively.

Response to general comments:

Thank you for your valuable feedback and suggestive comments on our manuscript, which not only recognizes the innovation and contribution of our approach but also highlights areas for further enhancement of our work. Below are our responses to your main points:

(1) The criteria for feature selection and the impact of varying combinations of observation sites

(1.1) The criteria for feature selection

The mean cross-validation score (MCV) is used as the criterion for feature selection, and the optimal feature subset is selected as the one that achieves the highest MCV. This selection is implemented by recursively removing the feature with the least importance, and assessing the MCV based on cross validation (Akhtar et al., 2019). It starts with training a XGBoost model with all features, and assessing the importance of each feature based on its contribution to the accuracy of the XGBoost model. Then, the feature with the least importance is removed and the XGBoost model is retrained using the remaining features. The feature importance and MCV are updated accordingly for the removal of another feature. This iteration continues until the optimal number of features is identified, corresponding to the highest MCV achieved during the process.

Using this criterion, unimportant features can be removed to improve the XGBoost model's prediction accuracy, while simultaneously reducing the risk of overfitting and computational costs.

To clearly reflect the criteria of feature selection, we have added some descriptions in relevant section.

► Lines 202-204 of section "2.5.3 Automatic optimization of XGBoost model" have been replaced with:

"*The initial input features (Table 1) are optimized through a feature selection step, which finds the optimal feature subset that achieves the best MCV. The selection is implemented by recursively removing the feature with the least importance,*

and reassessing the MCV based on cross validation (Akhtar et al., 2019). It starts with training a XGBoost model with all features, and assessing the importance of each feature based on its contribution to the model accuracy. The feature with the least importance is removed and the XGBoost model is retrained using the remaining features. The feature importance and MCV are updated accordingly for the removal of another feature. This iteration continues until the optimal number of features is identified, corresponding to the highest MCV achieved during the process. The overall flowchart of the proposed spatiotemporally decoupled source reconstruction model is shown in Fig. S1."

(1.2) Impact of varying combinations of observation sites

We agree that it is important to discuss the impact of varying combinations of observation sites. Briefly speaking, the selection of representative sites is more important for model performance than increasing the number of sites. In this study, we have demonstrated this impact through sensitivity studies with respect to both the number and combination of observation sites.

1) The number of observation sites

Additional observation sites can better capture environmental variability and impose stronger constraints on the estimation, leading to more robust results. However, the usage of all observation sites may cause overfitting of the XGBoost model and reduce the prediction accuracy of the trained model. Our sensitivity study (Fig. 12) also reveals that locating the source using all observation sites does not achieve the lowest error level, though the error level remains low.

2) The position of observation sites

Observation sites located at appropriate position can capture environmental variability and provide adequate information for locating the source. Utilizing only these representative sites can alleviate overfitting and enhance the prediction accuracy of the XGBoost model. The sensitivity study demonstrates that the lowest error levels are achieved by a subset of sites, i.e. Site ABD on Oct. 3 and Site BD on Oct. 4. For Oct.3, multi-site estimations with Site B always produce lower error levels, and single-site estimation using Site B also achieves high accuracy. For Oct.4, multi-site estimations with Site BD always achieve relatively low error levels. These results prove the importance of representative sites in source location estimation. In addition, the representative sites (Site B for Oct. 3 and Site BD for Oct. 4) are consistent with the feature selection results in Fig. 4, preliminarily indicating the potential of feature selection to identify representative sites.

To highlight the impact of varying combinations of observation sites, we have added some descriptions in relevant section.

► Lines 438-449 of section "3.4.4 Sensitivity to the number and combination of observation sites" have been replaced with:

"Figure 12 compares the results obtained with different numbers and combinations of observation sites. The results indicate that the localization error may be more sensitive to the position of the observation site than to the number of sites included. *The error level of all-site estimations is relatively low for both days, indicating that increasing the number of observation sites can better constrain the solution and help improve the robustness of the model. However, the lowest error levels are achieved by a subset of sites, i.e. Site ABD on Oct. 3 and Site BD on Oct. 4. This is possibly because the usage of all observation sites may cause overfitting and reduce the prediction accuracy of the XGBoost model. On the other hand, using only representative sites at appropriate position is effective to alleviate overfitting and enhance the model accuracy, because these representative sites can capture environmental variability and provide adequate information for locating the source. For Oct.3, multi-site estimations with Site B always produce lower error levels, and*

single-site estimation using Site B also achieves high accuracy. For Oct.4, multi-site estimations with Site BD always achieve relatively low error levels. These results prove the importance of representative sites in source location estimation. The representative sites (Site B for Oct. 3 and Site BD for Oct. 4) are consistent with the importance calculated in the feature selection (Fig. 4), preliminarily indicating the potential of feature selection to identify representative sites. In addition, feature selection also reduces the mean error level in most cases."

(2) More validation of the method:

We acknowledge the importance of validating our method against diverse scenarios and weather conditions to assess its robustness and practical applicability.

To address your concern, we have incorporated an additional validation case based on the first release of the European Tracer Experiment (ETEX-1) (Nodop et al., 1998), involving a different type of releases (continental-scale) and more complex meteorological conditions (temporally and spatially varying), to thoroughly assess the method's efficiency and adaptability. During ETEX-1, a total of 340 kg of perfluoromethylcyclohexane (PMCH) was released continuously into the atmosphere from 23 Oct 1994 16:00:00 UTC and 24 Oct 1994 03:50:00 UTC. Assuming the release could have occurred between 23 Oct 1994 00:00:00 UTC and 28 Oct 1994 00:00:00 UTC, it is viewed as a temporally-varying release, with a release rate of zero outside the actual release window. A total of 3104 available observations (3h-averaged concentrations) were collected by 168 sites. As shown in Fig. 2b, we choose two groups of observation sites: the first comprises four sites (i.e. B05, D10, D16, F02) randomly selected from the sites within the sample zone (Group1, with a total of 92 available observations), and the second involves four sites (i.e. CR02, D15, DK08, S09) randomly selected from the sites beyond the sample zone boundaries (Group2, with a total of 90 available observations).

For the continental-scale ETEX-1 experiment, the proposed method still achieves the lowest source location errors among all methods, which are below 10 km and 20 km (less than the grid size of 0.25°×0.25°) for Group1 and Group2, respectively. Regarding the results of the uncertainty analysis, the average relative source location error was 2.42% and 4.97% for Group1 and Group2, respectively, lower than the correlation-based and Bayesian methods. The proposed method provides highly accurate mean estimates of release rate for both groups, although with a large uncertainty range. These results demonstrate that spatiotemporally decoupled source reconstruction is feasible and achieves satisfactory accuracy in multi-scale release scenario, thereby providing a promising framework for reconstructing atmospheric radionuclide releases.

An overview of the ETEX-1 experiment and corresponding source reconstruction results are provided below. Detailed results and discussions have been included in the revised manuscript.

2.6 Multi-scale Validations

2.6.1 Field experiments

The ETEX-1 experiment took place at Monterfil in Brittany, France, on 23 October 1994 (Nodop et al., 1998). During ETEX-1, a total release of 340 kg of PMCH was released into the atmosphere during 23 Oct 1994 16:00:00 UTC and 24 Oct 1994 03:50:00 UTC. As illustrated in Fig. 2b, the source coordinates were (2.0083°W, 48.058°N). A total of 3104 available observations (3h-averaged concentrations) were collected by 168 sites. ETEX-1 has been commonly used as a validation scenario for reconstructing atmospheric radionuclide releases (Ulimoen and Klein, 2023; Tomas et al., 2021). The candidate source locations are uniformly sampled from the green shaded zone. We choose two groups of observation sites: the first comprises four sites (i.e. B05, D10, D16, F02) randomly selected from the sites within the sample zone (Group1, with a total of 92 available observations), and the second involves four sites (i.e. CR02, D15, DK08, S09) randomly selected from the sites beyond the sample zone boundaries (Group2, with a total of 90 available observations).

Compared to the SCK-CEN ^{41}Ar experiment, the observations of the ETEX-1 exhibit temporal sparsity, lower temporal resolution, and increased complexity in meteorological conditions.

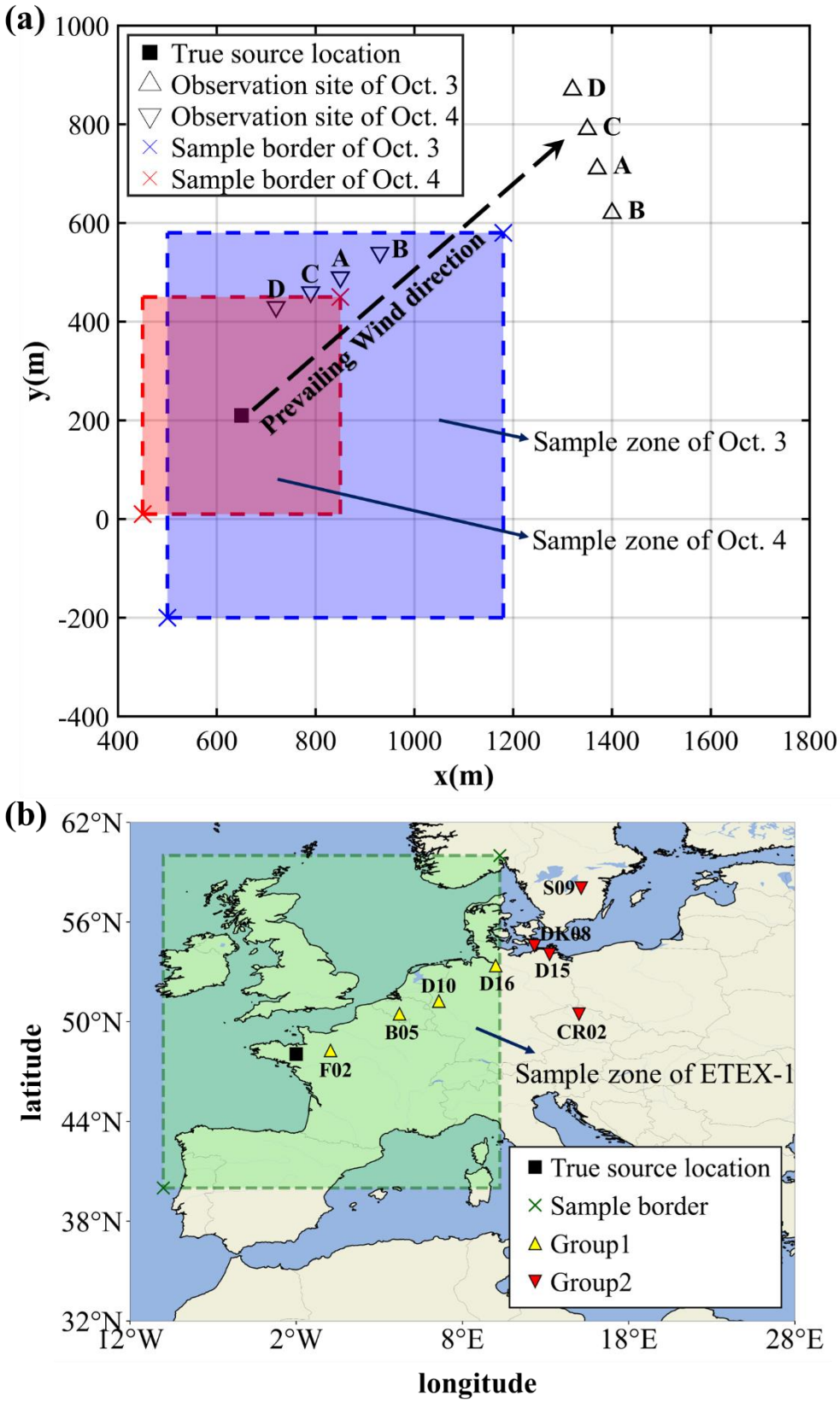


Figure 2. Release location and observation sites of two filed experiments. (a) SCK-CEN ^{41}Ar experiment. The map was created based on the relative positions of the release source and observation sites, as detailed in (Drews et al., 2002). The

coordinate of the sample border is (500m, -200m) and (1180m, 580m) on Oct. 3, and (450m, 10m) and (850m, 450m) on Oct. 4. It was plotted using MATLAB 2016b, instead of created by a map provider; (b) ETEX-1 experiment. The map was created based on the real longitudes and latitudes of the release source and observation sites, as detailed in (Nodop et al., 1998). The coordinate of the sample border is (10°W, 40°N) and (10°E, 60°N). It was plotted using the function *cartopy* of Python, instead of created by a map provider.

2.6.2 Simulation settings of atmospheric dispersion model

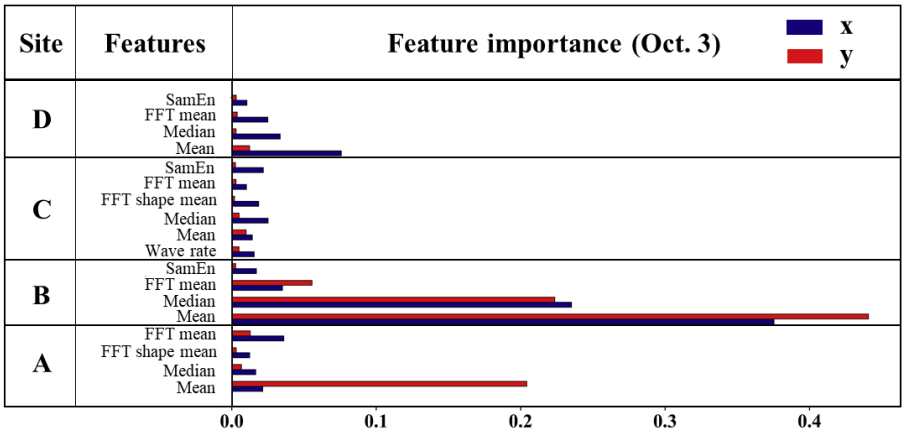
For the ETEX-1 experiment, the FLEXible PARTicle (FLEXPART) model (version 10.4) was applied to simulate the dispersion of PMCH (Pisso et al., 2019). The meteorological data were from the United States National Centers of Environmental Prediction Climate Forecast System Reanalysis, which has a spatial resolution of 0.5°×0.5° and time resolution of 6h. To rapidly establish the relationship between the varying source locations and the observations, 182 backward simulations were performed using FLEXPART. The candidate source locations were uniformly sampled from the shaded zone in Fig. 2b, resulting a total of 6561 source locations. As described in Sect. 2.5.1, 2624 candidate source locations are preserved by pre-screening step. The constant factors mentioned in Sect. 2.5.2 are 5.60×10^{12} and 2.86×10^{13}.

3.2 Optimization of XGBoost model

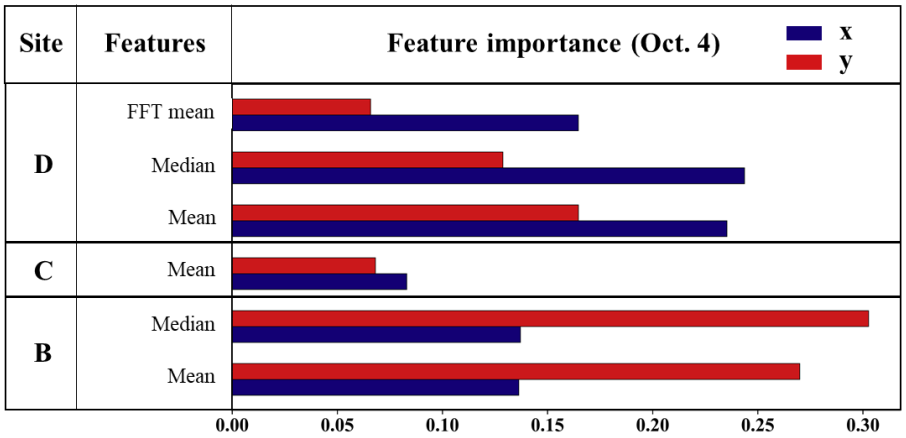
3.2.2 Feature selection

Figure 4 compares the importance of the selected features at each site. For the ETEX-1 experiment, Fig. 4c and d showed that the features of Group1 and Group2 are almost reserved after the feature selection process (only one feature is removed for each case), indicating fewer redundant features than that in the SCK-CEN ^{41}Ar experiment. The time-domain features are also dominant, but the frequency-domain features at some sites also play important roles such as D16 and S09. The MCVs of the ETEX-1 experiment have similar variation trend as that of the SCK-CEN ^{41}Ar experiment (Fig. S4c and d). However, the Group1 has obviously lower MCV than other experiments.

(a)


Site
Features
Feature importance (Oct. 3)
x
y
D
SamEn
FFT mean
Median
Mean
C
SamEn
FFT mean
FFT shape mean
Median
Mean
Wave rate
B
SamEn
FFT mean
Median
Mean
A
FFT mean
FFT shape mean
Median
Mean
0.0
0.1
0.2
0.3
0.4

(b)


Site
Features
Feature importance (Oct. 4)
x
y
D
FFT mean
Median
Mean
C
Mean
B
Median
Mean
0.00
0.05
0.10
0.15
0.20
0.25
0.30

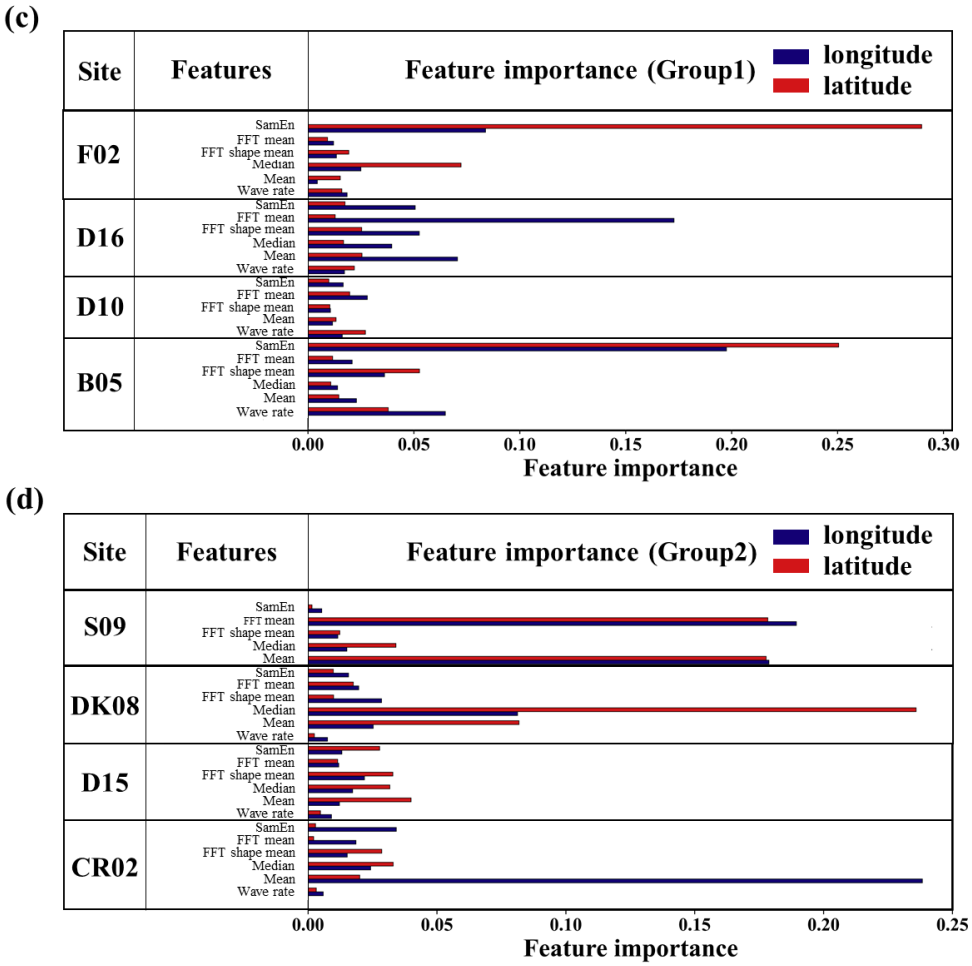


Figure 4. Feature importance. (a) Oct. 3; (b) Oct. 4; (c) Group1; (d) Group2.

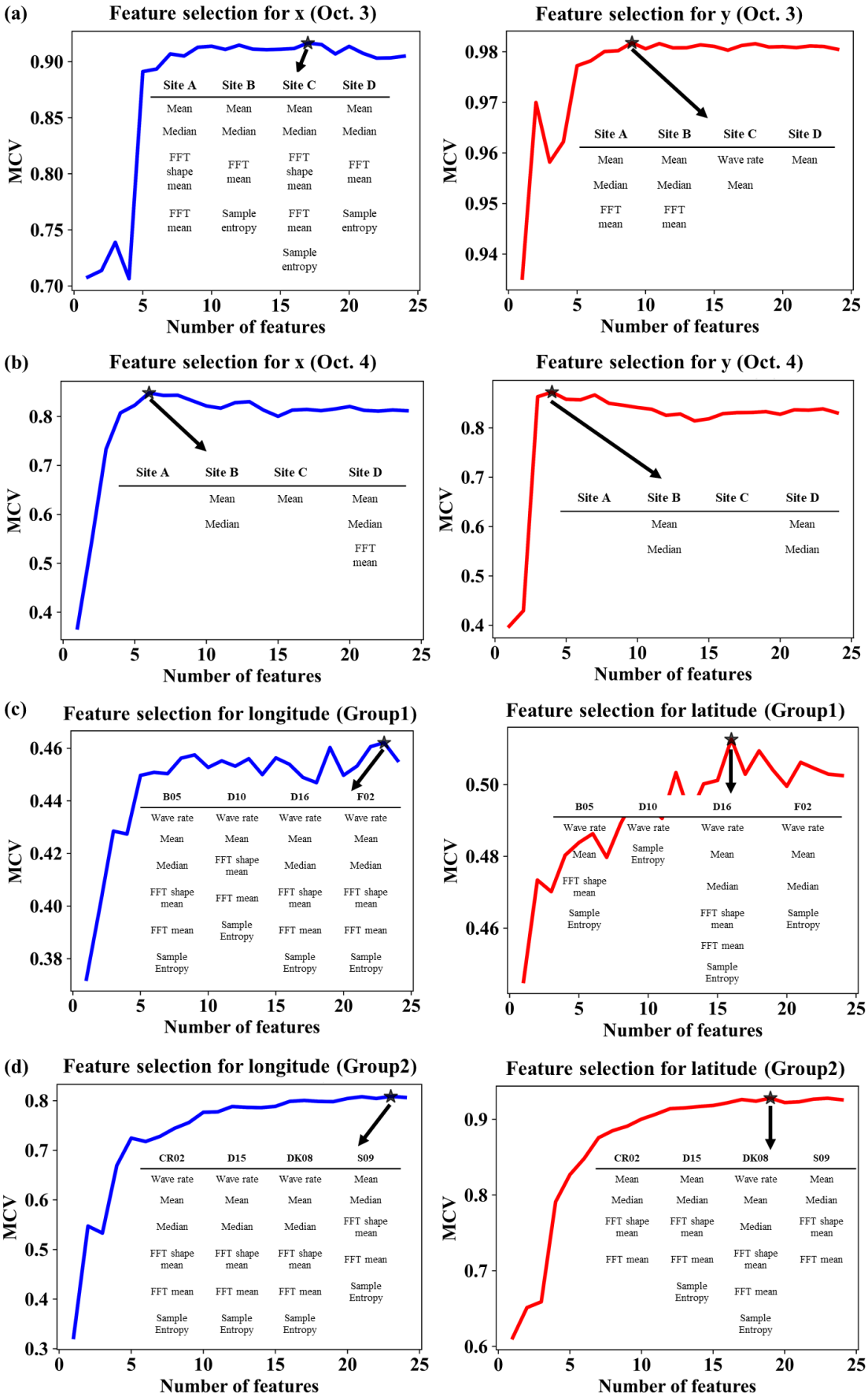


Figure S4. Results of feature selection in x (longitude) and y (latitude) directions. (a) Oct. 3; (b) Oct. 4; (c) Group1; (d) Group2. The black stars denote the optimal number of features. The table inserted in each subgraph lists the selected features for each observation site.

3.3 Source reconstruction

3.3.1 Source location estimation

Figure 5 compares the best-estimated source locations of the correlation-based method, the Bayesian method, and the proposed method with the ground truth. For the ETEX-1 experiment, the pre-screening zone also covers the true source location for Group1 and Group2. But source locations estimated by the correlation-based method are 411.85 km and 486.41 km away from the ground truth for Group1 and Group2, respectively. The location error of the Bayesian method estimates is only 30.50 km for Group1, but it increases to 520.77 km for Group2, indicating its sensitivity to the observations. In contrary, the proposed method achieves the lowest source location errors, which are below 10 km and 20 km for Group1 and Group2, respectively. Because the feature selection hardly removed the features (Fig. 4c and d), the estimated source locations with and without feature selection basically overlap for both groups.

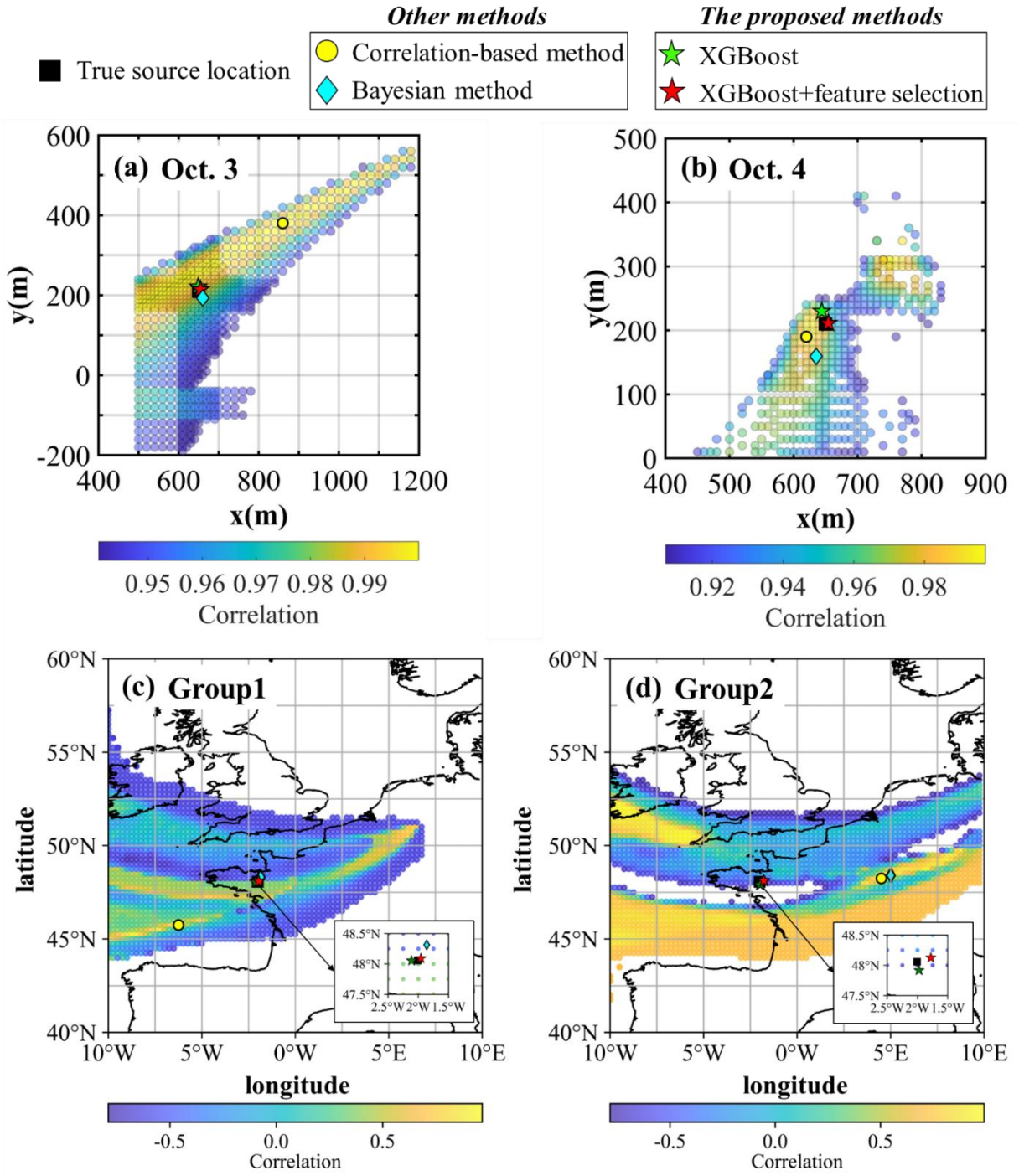


Figure 5. Source location estimation results. The yellow dots denote the maximum correlation points, which are the localization results of the correlation-based method. The green and red stars represent the localization results based on XGBoost before and after feature selection, respectively. The cyan diamonds represent the localization results based on the Bayesian method. (a) Oct. 3; (b) Oct. 4; (c) Group1; (d) Group2. A detailed enlargement of the region around (2.5°W, 47.5°N) to (1.5°W, 48.5°N) is shown in the bottom right corner in (c) and (d) to highlight the source location estimation results of the proposed methods.

3.3.2 Release rates

Figure 6 displays the release rates estimated by the Bayesian and PAMILT methods based on the source location estimates in Fig. 5. For the two ETEX-1 groups (Fig. 6c and d), Bayesian estimates exhibit noticeable fluctuations, leading to underestimations–58.11% for Group1 and 51.44% for Group2. Furthermore, the temporal profile of the Bayesian estimates for Group2 completely falls out of the true release window. In contrast, most releases of the PAMILT estimates are within the true release time window, especially for Group2, despite overestimations reach 52.38% for Group1 and 57.65% for Group2, after feature selection process. Compared to the SCK-CEN ^{41}Ar experiment, the increased deviation in the ETEX-1 experiment may be attributed to the limited observations at the four sites as shown in Fig. S3.

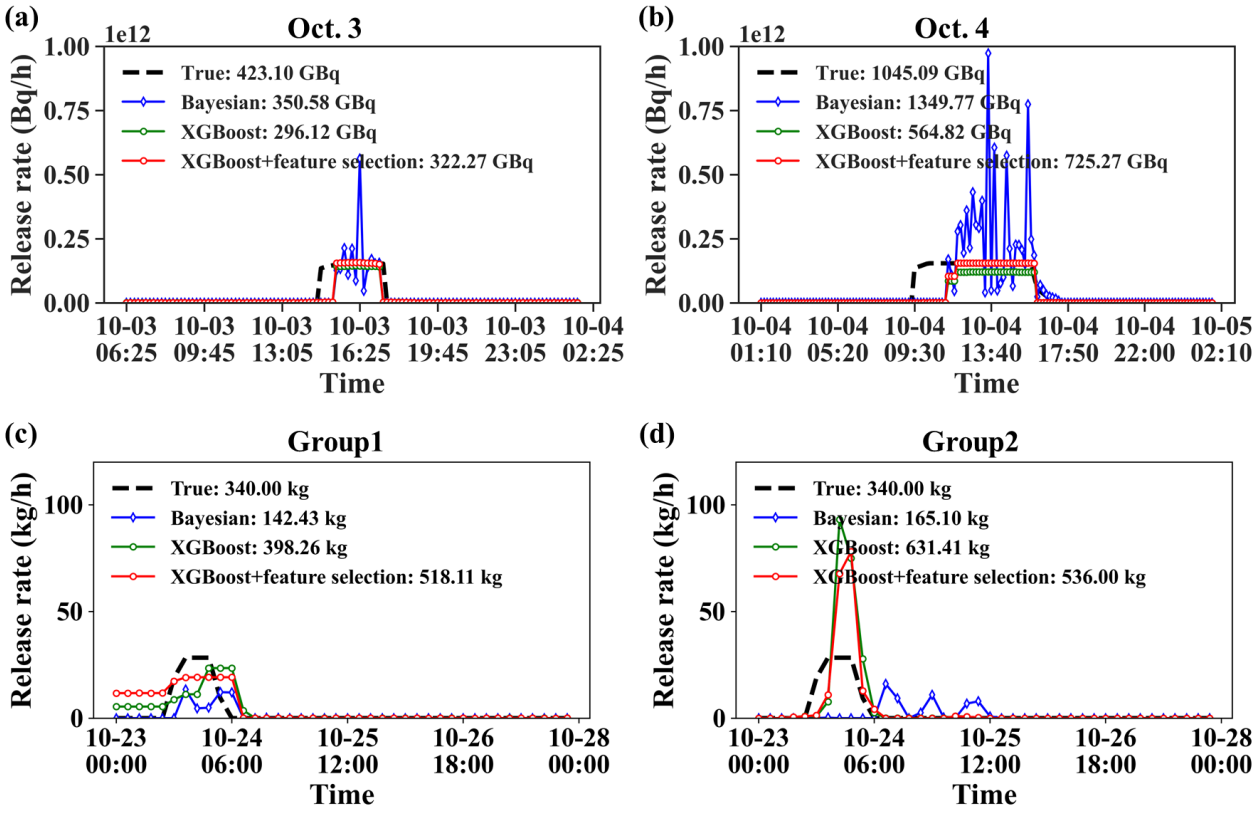


Figure 6. Release rate estimation results with different location estimates. (a) Oct. 3; (b) Oct. 4; (c) Group1; (d) Group2. The release rates labelled by XGBoost or XGBoost+feature selection are estimated using PAMILT method.

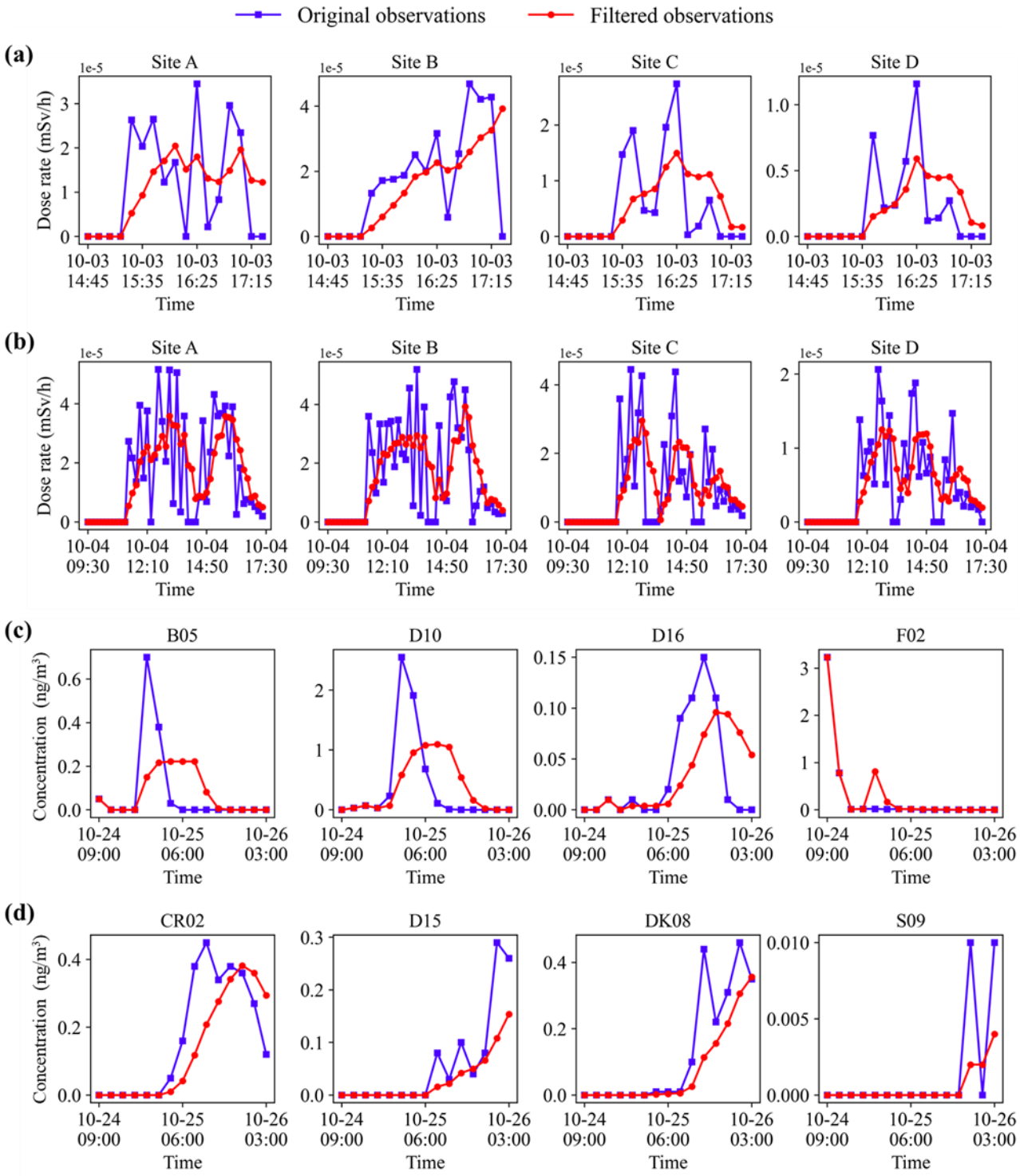


Figure S3. Observations before and after filtering at observation sites. (a) Oct. 3; (b) Oct. 4; (c) Group1; (d) Group2.

3.3.3 Uncertainty range

Figure 7 compares the spatial distribution of 50 estimates produced by the different source location estimation methods. For the ETEX-1 experiment, the estimates of the correlation-based method are quite spread, whereas those of the Bayesian method are more concentrated. The Bayesian estimates are close to the truth in Group1, but are noticeably deviated in Group2. The phenomenon indicates that the Bayesian method is sensitive to the observations, especially when observations are limited. Fig. S5c and d also reveals that the Bayesian-estimated posterior distribution is multimodal for both ETEX-1 groups, which can be avoided by using additional observations (Fig. S5e). In contrast, the

proposed method provides estimates quite concentrated around the truth for both Group1 and Group2, indicating its efficiency in the case of limited observations. Compared to Group2, Group1 exhibits lower source location error. This is attributed to the fact that, as shown in Fig. 2, Group1's four observation sites are situated closer to the sampled source locations than those of Group2, thereby providing more accurate plume characteristics and resulting in lower dispersion error.

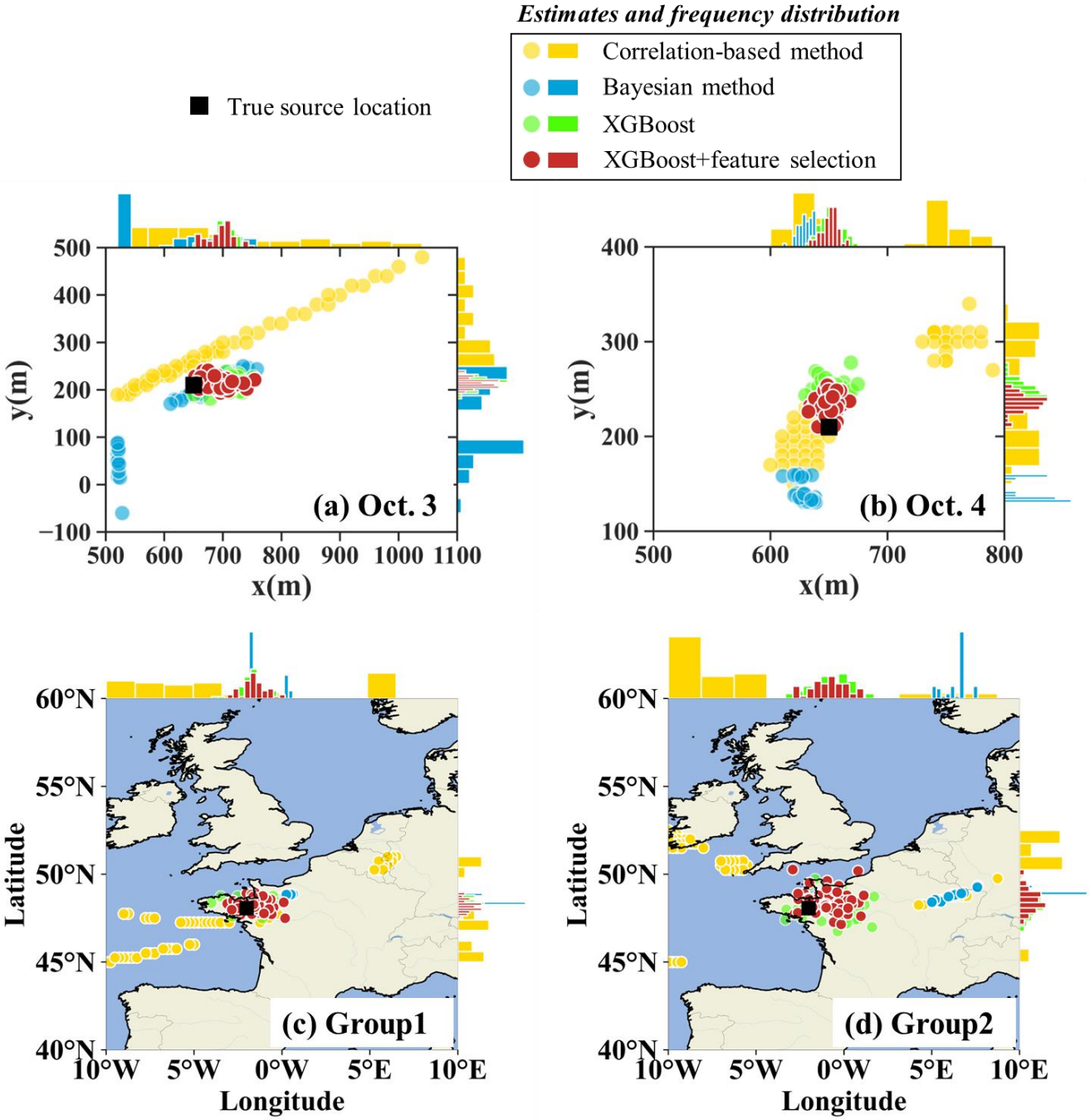


Figure 7. Spatial distribution of 50 source location estimates. (a) Oct. 3; (b) Oct. 4; (c) Group1; (d) Group2. Each circle denotes an individual estimate as detailed in Sect. 2.8.5, with color variations indicating the respective method employed. Histograms along the axes represent the frequency distribution of the estimates along the respective axis.

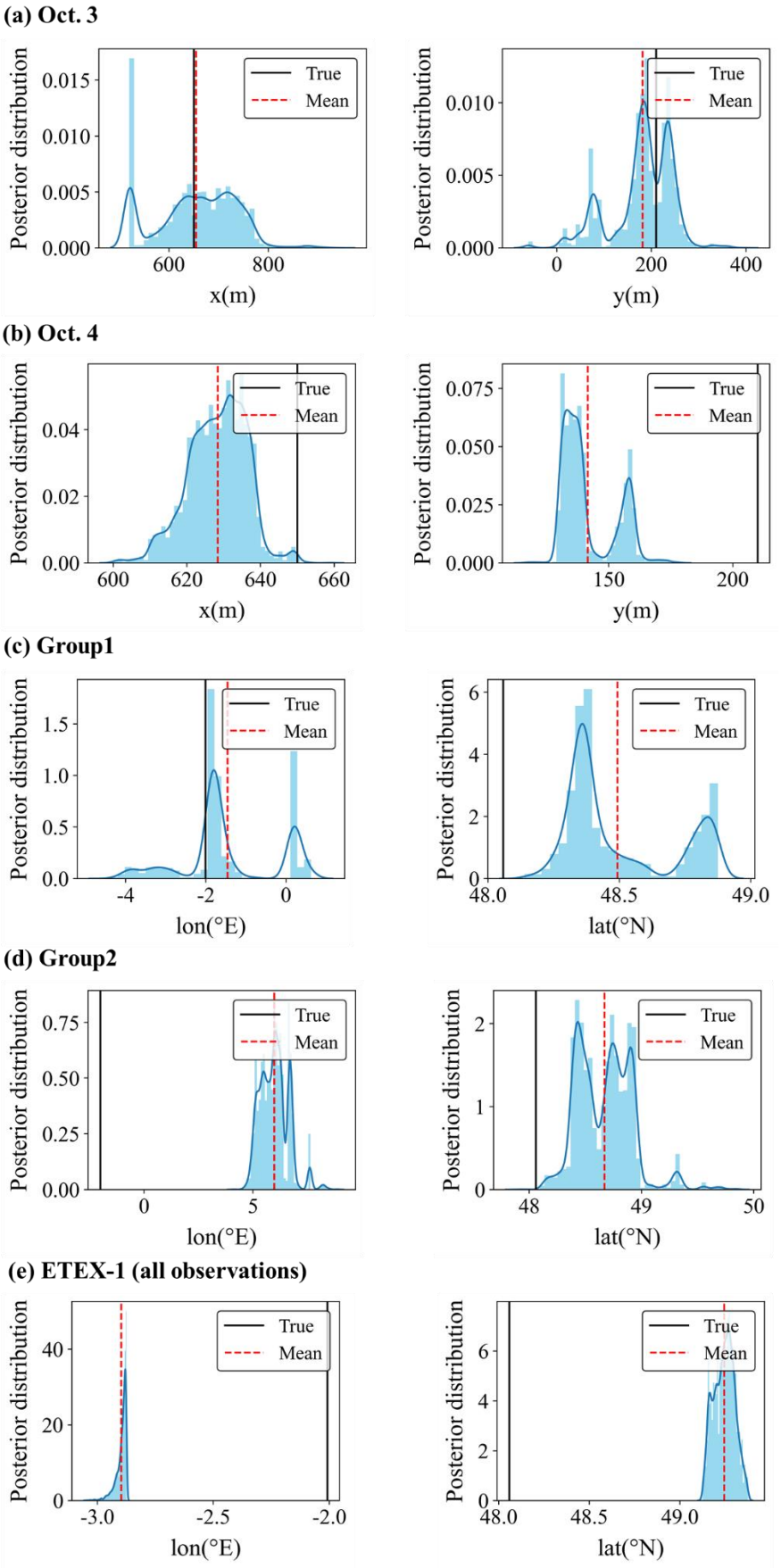


Figure S5. Posterior distributions of source location parameters. (a) Oct. 3; (b) Oct. 4; (c) Group1; (d) Group2; (e) ETEX-1 (all observations in ETEX-1 are used). The black solid lines denote the true location parameters and the dashed lines denote the mean estimates of all posterior samples.

Figure 9 compares the uncertainty ranges of the release rate estimates in the two ETEX-1 groups. For both groups, the Bayesian estimates show underestimations (including the mean estimate), but it presents small uncertainty range (Fig. 9a and c). In addition, the Bayesian estimates all fall out of the time window of the true release in Group2 (Fig.9c). The mean PAMILT estimates are more accurate mean Bayesian estimates, which most releases are within the time window of the true release (Fig. 9b and d). However, PAMILT estimates for the ETEX-1 experiment show a large uncertainty range in both groups, in contrast to the SCK-CEN ^{41}Ar experiment. The reason for this discrepancy is that the source-receptor matrices of the ETEX-1 experiment are more sensitive to errors in source location than those of the SCK-CEN ^{41}Ar experiment. The greater sensitivity originates from the ETEX-1 experiment's more complex meteorological conditions, both spatially and temporally. As for the mean total release, the Bayesian method shows underestimations of 70.93% for Group1 and 74.15% for Group2. In comparison, the proposed method shows deviation of only 0.71% for Group1 and 0.09% for Group2, after feature selection.

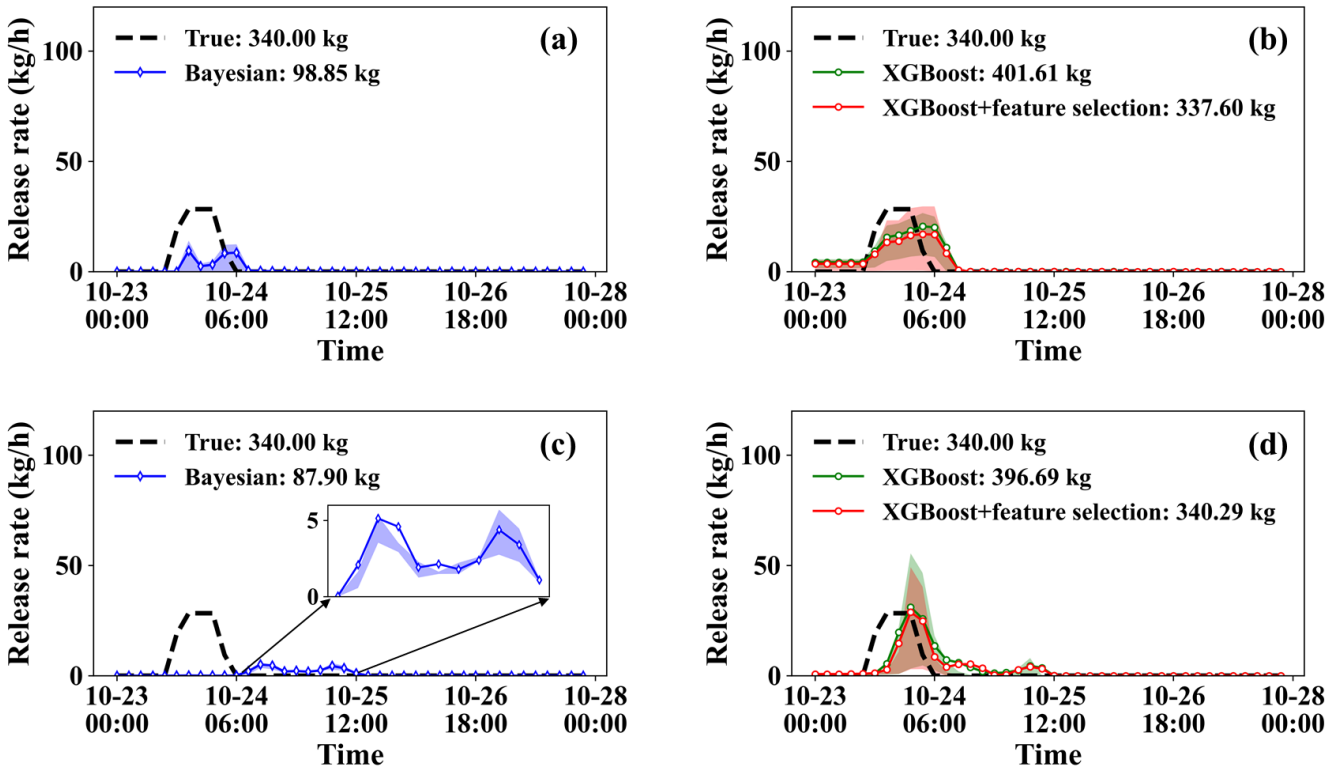


Figure 9. Release rate estimates over 50 calculations in ETEX-1 experiment. (a) Group1-Bayesian method; (b) Group1-PAMILT method; (c) Group2-Bayesian method; (d) Group2-PAMILT method.

Table 3 lists the mean and standard deviation of the relative errors for the 50 estimates given by different methods. For the ETEX-1 experiment, the Bayesian method shows case-sensitive performances with respect to the mean relative error of source location estimation, whereas the proposed method gives the most accurate source locations with small uncertainties for both groups. As for the total release, the proposed methods show smaller mean relative errors than the Bayesian methods, but the Bayesian method shows a smaller standard deviation.

Table 3. Relative errors of source reconstruction. $\delta_{\mathbf{r}}$ represents the relative error of source location, which is positive and δ_Q denotes the relative error of total release, where a positive value indicates overestimation and a negative value denotes underestimation.

Experiments	Statistical parameters (Relative error)		Correlation-based method	Bayesian method	The proposed method	
					XGBoost	XGBoost+ feature selection
Oct. 3	$\delta_{\mathbf{r}}$	Mean	14.10%	11.88%	5.18%	4.68%
		Std	11.37%	7.53%	1.79%	2.05%
	δ_Q	Mean	-	153.61%	-16.93%	-18.30%
		Std	-	189.76%	9.45%	8.01%
Oct. 4	$\delta_{\mathbf{r}}$	Mean	14.30%	12.83%	6.83%	4.71%
		Std	9.60%	1.68%	1.76%	1.53%
	δ_Q	Mean	-	42.29%	-54.12%	-47.42%
		Std	-	15.05%	6.47%	5.85%
Group1	$\delta_{\mathbf{r}}$	Mean	16.95%	3.22%	2.32%	2.42%
		Std	7.46%	2.75%	1.43%	1.43%
	δ_Q	Mean	-	-70.93%	18.12%	-0.71%
		Std	-	17.87%	99.85%	102.01%
Group2	$\delta_{\mathbf{r}}$	Mean	21.9%	23.97%	5.21%	4.97%
		Std	5.05%	1.97%	2.42%	2.35%
	δ_Q	Mean	-	-74.15%	16.67%	0.09%
		Std	-	11.68%	93.50%	109.56%

Specific comments

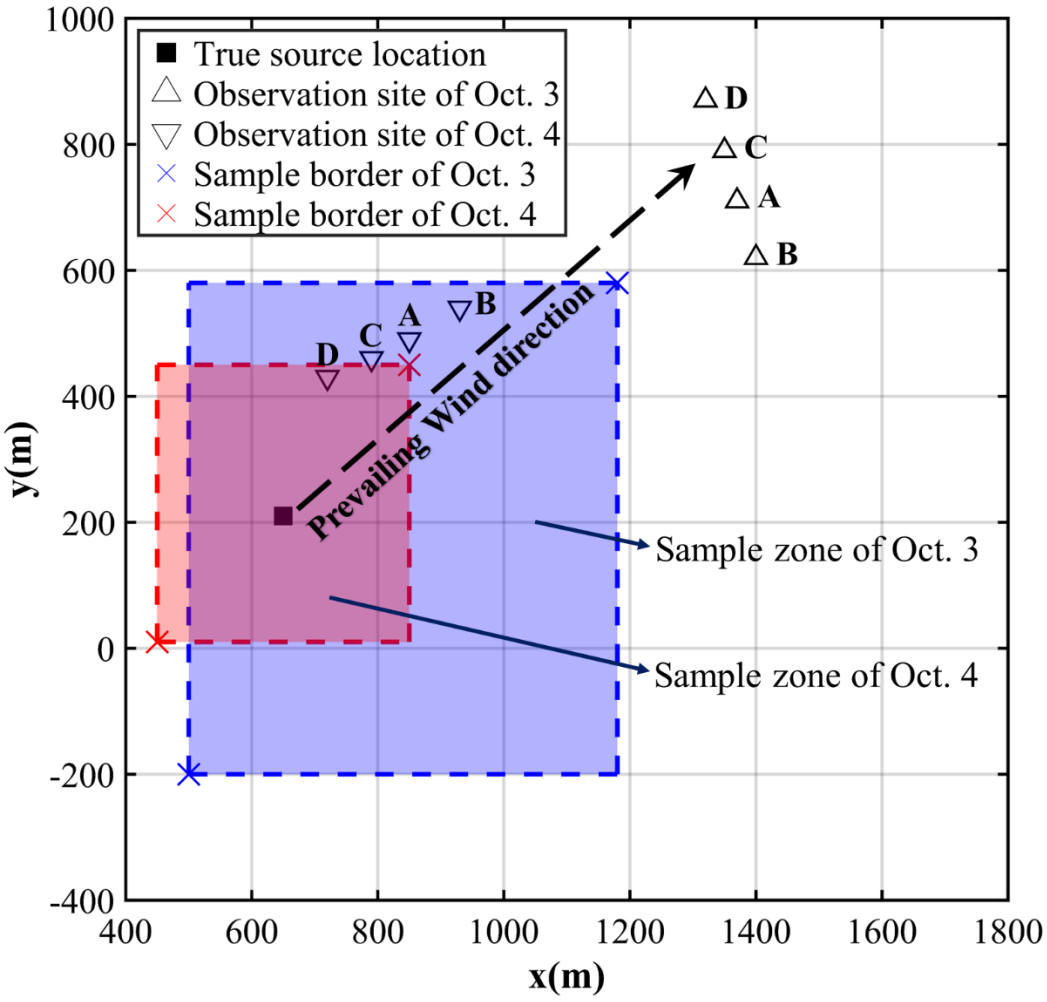
Comment#1:

L215 - Figure 2: The axes represent distances and should therefore have identical scales for clarity and accuracy.

Response to comment#1:

We appreciate your attention on the different scales of axes on Figure 2. We have revised the figure to ensure that both axes represent distances with identical scales.

► Figure 2 has been replaced with:



"**Figure 2**. Release location and observation sites of SCK-CEN ^{41}Ar experiment. The map was created based on the relative positions of the release source and observation sites, as detailed in (Drews et al., 2002). It was plotted using MATLAB 2016b, instead of created by a map provider."

Comment#2:

L222 - Consideration of vertical information could provide a more comprehensive understanding of the dispersion patterns. How does the model account for vertical dispersion?

Response to comment#2:

Thank you for your constructive comment. We agree that incorporating vertical information can aid in understanding the dispersion patterns. In the SCK-CEN ^{41}Ar experiment, the ^{41}Ar was emitted from a 60-m stack, while the ground-level fluence rates were collected by NaI (Tl) gamma detectors. Due to the lack of vertical observations in

this experiment, the vertical dispersion has not been discussed in the manuscript.

The RIMPUFF model is a gaussian puff model that uses the diffusion coefficient in vertical direction to describe the vertical dispersion of each puff. In this study, the Karlsruhe-Jülich diffusion coefficients were used to calculate the vertical dispersion, which has been validated for the SCK-CEN ^{41}Ar experiment and has shown good accuracy (Li et al., 2019).

In the future, we will try to incorporate the vertical dispersion information into the source parameters.

To ensure clarity, we have added some descriptions of vertical information in the revised manuscript.

► Lines 218 of section "2.6.1 SCK-CEN ^{41}Ar field experiment" have been replaced with:

"The 60-s-average *ground-level* fluence rates were continuously collected by an array of NaI (Tl) gamma detectors,"

► Lines 225-226 of section "2.6.2 Simulation settings of atmospheric dispersion model" have been replaced with:

"The calculation domain measured 1800 m×1800 and the grid resolution was 10 m×10 m. *The release height of ^{41}Ar was presumed to be 60 m.*"

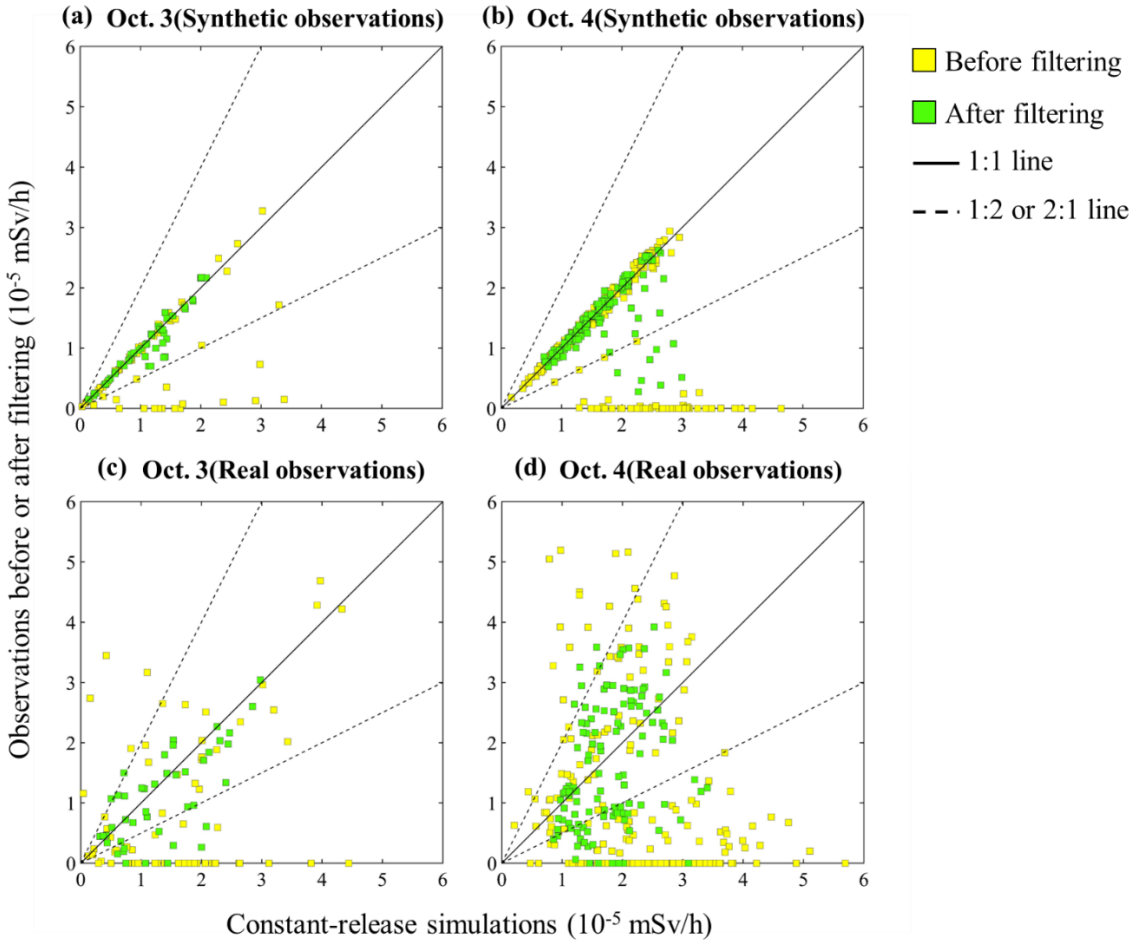
Comment#3:

L302 - Figure 3: To ensure clarity and accuracy in data representation, the scales on the vertical and horizontal axes must be consistent.

Response to comment#3:

We appreciate your attention on the different scales on axes on Figure 3. We have carefully adjusted Figure 3 so that the scales on both the vertical and horizontal axes are now consistent.

► Figure 3 has been replaced with:



"**Figure 3**. Scatter plots of the original (yellow squares) and decoupled (green squares) observations versus the constant-release simulation results. (a) Oct. 3-Synthetic observations; (b) Oct. 4-Synthetic observations; (c) Oct. 3-Real observations; (d) Oct. 4-Real observations."

Comment#4:

L340 - The capability to estimate with greater accuracy than the grid size warrants a discussion. What implications does this have for the model's precision and its practical significance?

Response to comment#4:

Thank you for your insightful query. The ability to estimate source locations with accuracy surpassing the grid size can be attributed to the strong fitting capability of the optimized XGBoost model (Chen and Guestrin, 2016; Grinsztajn et al., 2022), which excels in interpolating within the grid size and extrapolating beyond the source location samples. As discussed in our response to the General comments, ETEX-1 experiment also achieved source location accuracy beyond the grid size (Fig. 5c and d, the grid size is 0.25°×0.25°), suggesting that the phenomenon is not merely coincidental. In addition, previous studies also achieved similar source location accuracy using traditional methods (Lucas et al., 2017; Tichý et al., 2017). However, this ability, although inherent, does not uniformly manifest across all optimized XGBoost models, as external factors like observation noises and meteorological data inaccuracies can also impact the accuracy of source location estimation. The uncertainty analysis in Sect. 3.3.3 has demonstrated that the source location estimates tend to cluster within several grids surrounding the true source, which is more reasonable and practical in real-world scenarios. Detailed discussions are as follows:

(1) ***Enhanced accuracy through XGBoost***: The high accuracy in locating the source is directly achieved by the XGBoost model, since it establishes the complex nonlinear relationships between the input features and the source location. Utilizing automatic optimization techniques (detailed in Sect. 2.5.3), 8 main hyperparameters of XGBoost and 24 observation series features are finely tuned to achieve an optimized model. This optimization not only mitigates the risk of overfitting but also enhances the model's ability for interpolation within the grid size and extrapolation beyond the source location samples.

(2) ***Validation on ETEX-1 experiment***: The proposed method has been validated through ETEX-1 experiment, as discussed in our response to the General comments. Compared to SCK-CEN ^{41}Ar experiment, ETEX-1 involves a different type of releases (continental-scale) and more complex meteorological conditions (temporally and spatially varying). As shown in Fig. 5c and d, this experiment also achieved source location accuracy beyond the grid size (0.25°×0.25°), suggesting that the phenomenon is not merely coincidental. Providing that the XGBoost model is effectively optimized and the observations are reliable, the model has ability to achieve high accuracy.

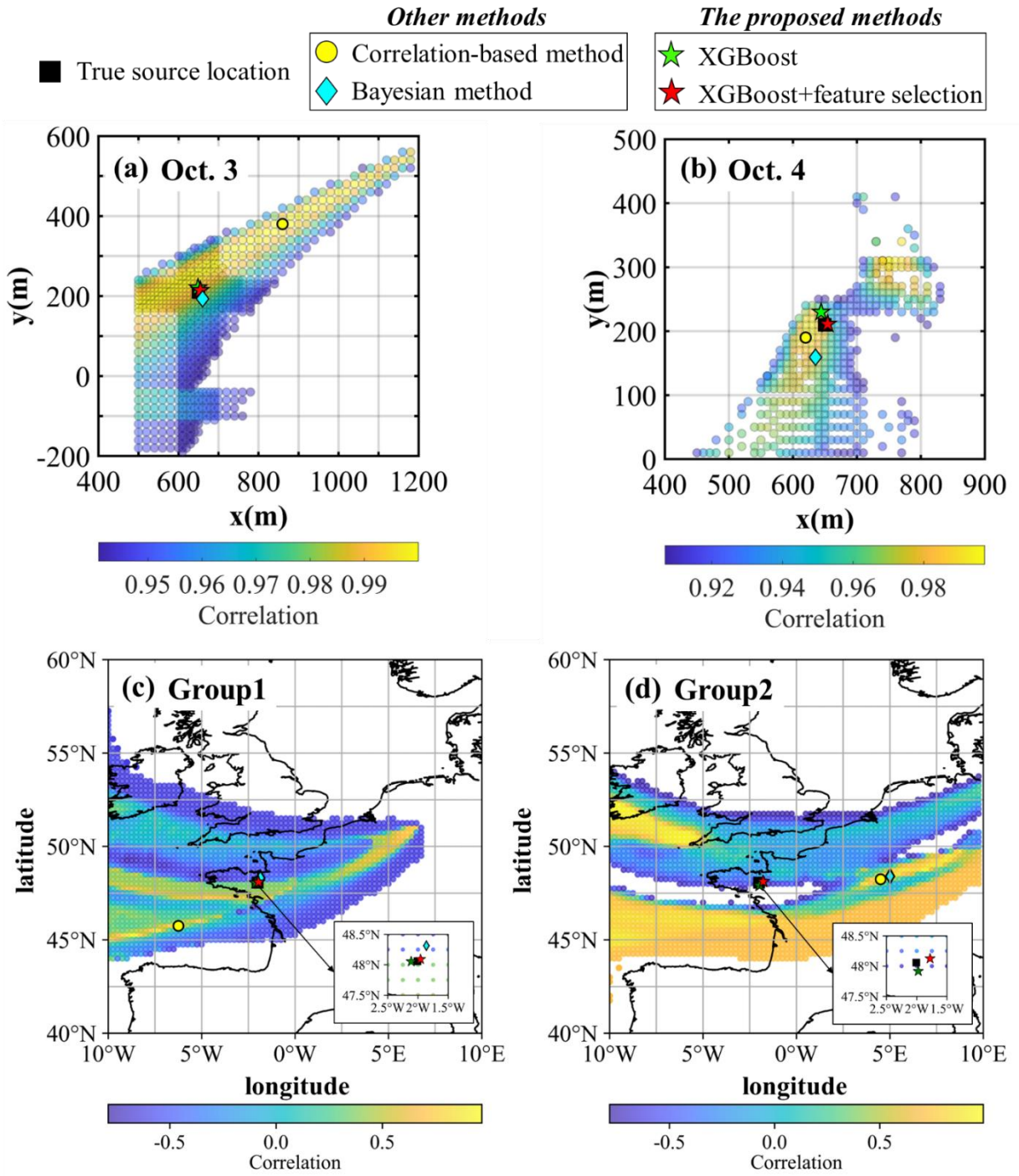


Figure 5. Source location estimation results. The yellow dots denote the maximum correlation points, which are the localization results of the correlation-based method. The green and red stars represent the localization results based on XGBoost before and after feature selection, respectively. The cyan diamonds represent the localization results based on the Bayesian method. (a) Oct. 3; (b) Oct. 4; (c) Group1; (d) Group2. A detailed enlargement of the region around (2.5°W, 47.5°N) to (1.5°W, 48.5°N) is shown in the bottom right corner in (c) and (d) to highlight the source location estimation results of the proposed methods.

(3) ***Uncertainty analysis of XGBoost hyperparameters***: An uncertainty analysis of the XGBoost hyperparameters (Fig. 7) has revealed that not all source location estimates achieve greater accuracy than the grid size. Instead, source location estimates tend to cluster within several grids surrounding the true source. This phenomenon highlights the practical significance of the proposed method.

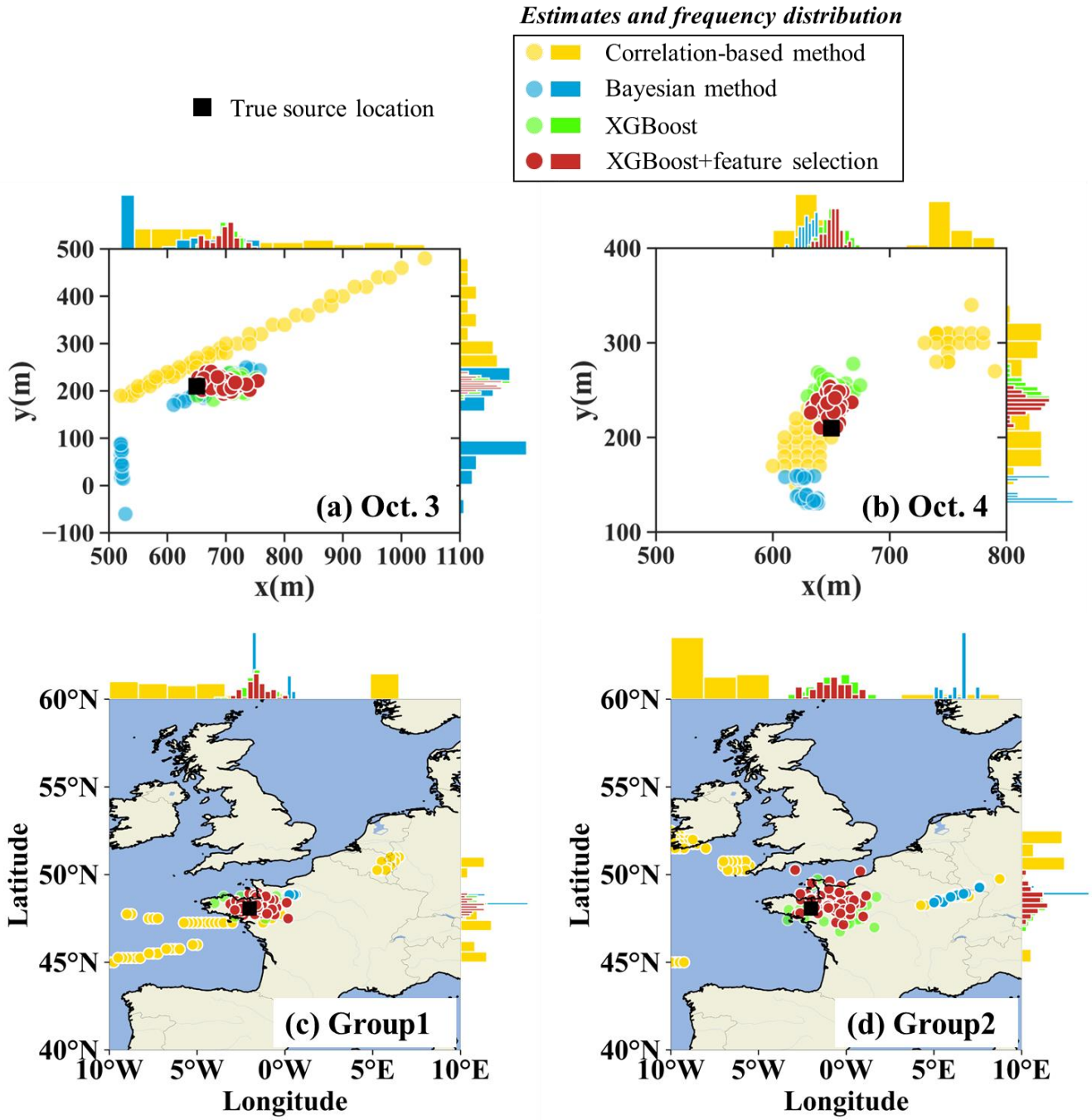


Figure 7. Spatial distribution of 50 source location estimates. (a) Oct. 3; (b) Oct. 4; (c) Group1; (d) Group2. Each circle denotes an individual estimate as detailed in Sect. 2.8.5, with color variations indicating the respective method employed. Histograms along the axes represent the frequency distribution of the estimates along the respective axis.

To avoid confusion, we have added some discussions in the revised manuscript to explain the greater accuracy than the grid size. The revised manuscript has also included the reconstruction results of ETEX-1 experiment (see our response to the General comments), which will further prove the model's ability.

► Lines 339-344 of section "3.3.1 Localization" have been replaced with:

"The estimates without feature selection are only 10.65 m (Oct. 3) and 20.62 m (Oct. 4) away from the true locations. Feature selection further reduces these errors to 6.19 m (Oct. 3) and 4.52 m (Oct. 4), which are below the grid size (10 m × 10 m) of the ATDM simulation. *The ability to estimate source locations with accuracy surpassing the grid size can be attributed to the strong fitting capability of the optimized XGBoost model* (Chen and Guestrin, 2016; Grinsztajn et al.,

2022), *which excels in interpolating within the grid size and extrapolating beyond the source location samples. However, this capability, although inherent, does not uniformly manifest across all optimized XGBoost models, as external factors like observation noises and meteorological data inaccuracies can also impact the accuracy of source location estimation. Therefore, uncertainty analysis will be performed in Sect. 3.3.3 to demonstrate the model's uncertainty range in estimating the source location.* The proposed method gives a relative error of less than 0.9% for both days, whereas the Bayesian method produces a relative error of above 11% and that of the correlation-based method can be as high as 26%. The best estimates for Oct. 3 are more accurate than those for Oct. 4, possibly because of the better layout of observation sites (Fig. 2) and the better decoupling results (Fig. 3)."

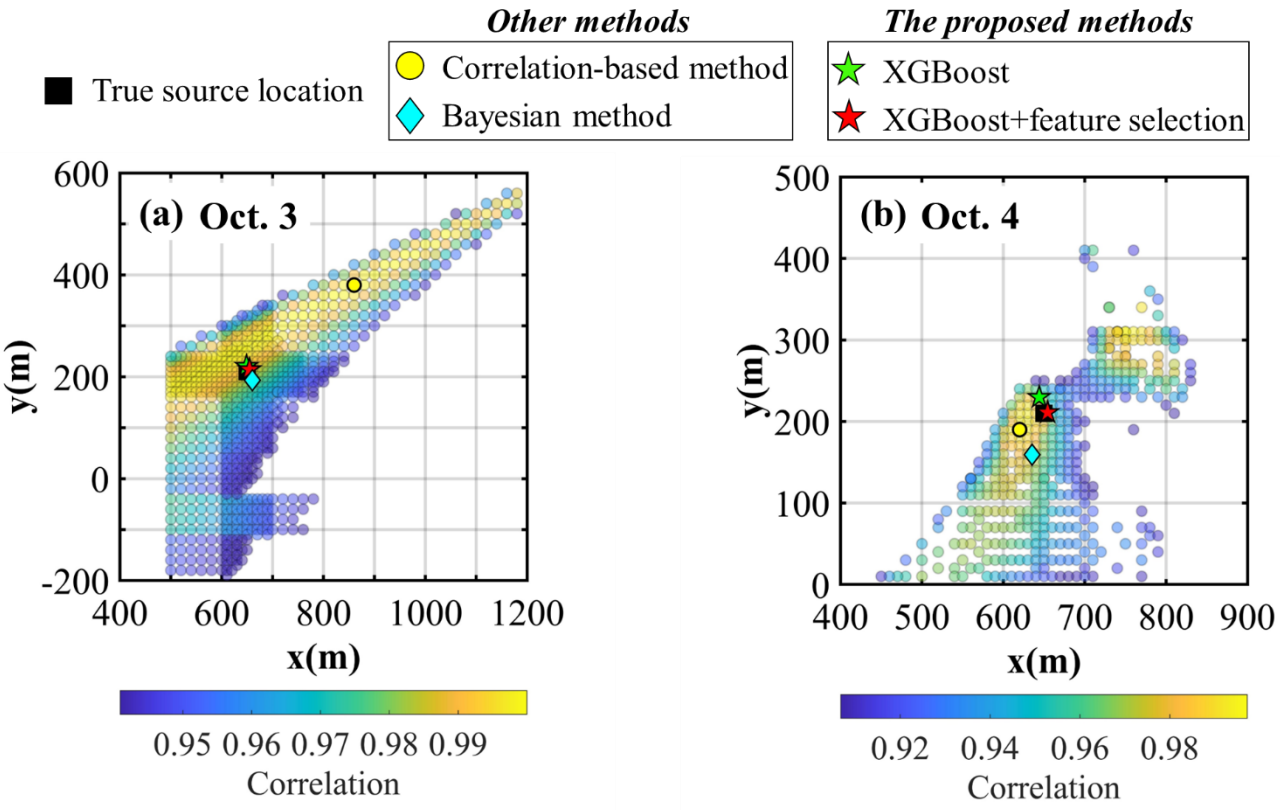
Comment#5:

L346 - Figure 5: As these axes represent distances, maintaining identical scales on both axes is crucial for accurate data interpretation.

Response to comment#5:

Thank you for pointing out this issue. We have revised Figure 5 to ensure that both the horizontal and vertical axes have the same scale.

► Figure 5 has been replaced with:



"**Figure 5.** Source localization results. The yellow dots denote the maximum correlation points, which are the localization results of the correlation-based method. The green and red stars represent the localization results based on XGBoost before and after feature selection, respectively. The cyan diamonds represent the localization results based on the Bayesian method. (a) Oct. 3; (b) Oct. 4."

Comment#6:

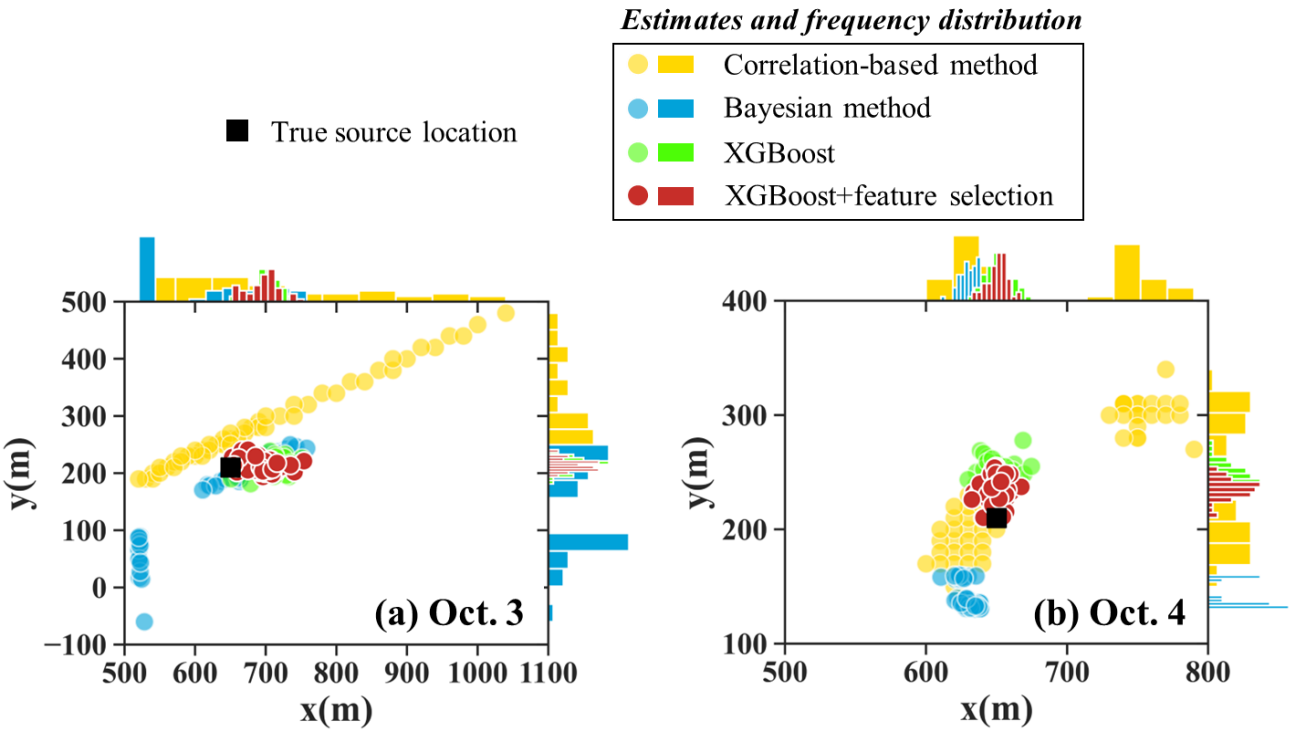
L374 - Figure 7: Given that both axes represent distances, their scales should be identical. The complexity of the graphs

necessitates a detailed explanation within the figure caption to aid in interpretation.

Response to comment#6:

Thank you for your constructive comment. We have revised Figure 7 to ensure that both axes are now on identical scales and expanded the figure caption to include a detailed explanation of the graph's components.

► Figure 7 and its figure caption have been replaced with:



"**Figure 7.** *Spatial distribution of 50 source location estimates. (a) Oct. 3; (b) Oct. 4. Each circle denotes an individual estimate as detailed in Sect. 2.8.5, with color variations indicating the respective method employed. Histograms along the axes represent the frequency distribution of the estimates along the respective axis.*"

Thanks again for such a thorough review!

References

Akhtar, F., Li, J., Pei, Y., Xu, Y., Rajput, A., and Wang, Q.: Optimal Features Subset Selection for Large for Gestational Age Classification Using GridSearch Based Recursive Feature Elimination with Cross-Validation Scheme, in: International Conference on Frontier Computing, 63–71, https://doi.org/10.1007/978-981-15-3250-4_8, 2019.

Chen, T. and Guestrin, C.: XGBoost: A scalable tree boosting system, Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 13-17-Augu, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Drews, M., Aage, H. K., Bargholz, K., Ejsing Jørgensen, H., Korsbech, U., Lauritzen, B., Mikkelsen, T., Rojas-Palma, C., and Ammel, R. Van: Measurements of plume geometry and argon-41 radiation field at the BR1 reactor in Mol, Belgium, 1–43 pp., 2002.

Grinsztajn, L., Oyallon, E., and Varoquaux, G.: Why do tree-based models still outperform deep learning on tabular data?, 2022.

Li, X., Xiong, W., Hu, X., Sun, S., Li, H., Yang, X., Zhang, Q., Nibart, M., Albergel, A., and Fang, S.: An accurate and ultrafast method for estimating three-dimensional radiological dose rate fields from arbitrary atmospheric radionuclide distributions, Atmos. Environ., 199, 143–154, https://doi.org/10.1016/j.atmosenv.2018.11.001, 2019.

Lucas, D. D., Simpson, M., Cameron-Smith, P., and Baskett, R. L.: Bayesian inverse modeling of the atmospheric transport and emissions of a controlled tracer release from a nuclear power plant, Atmos. Chem. Phys., 17, 13521–13543, https://doi.org/10.5194/acp-17-13521-2017, 2017.

Nodop, K., Connolly, R., and Girardi, F.: The field campaigns of the European tracer experiment (ETEX): Overview and results, Atmos. Environ., 32, 4095–4108, https://doi.org/10.1016/S1352-2310(98)00190-3, 1998.

Pisso, I., Sollum, E., Grythe, H., Kristiansen, N. I., Cassiani, M., Eckhardt, S., Arnold, D., Morton, D., Thompson, R. L., Groot Zwaaftink, C. D., Evangeliou, N., Sodemann, H., Haimberger, L., Henne, S., Brunner, D., Burkhart, J. F., Fouilloux, A., Brioude, J., Philipp, A., Seibert, P., and Stohl, A.: The Lagrangian particle dispersion model FLEXPART version 10.4, Geosci. Model Dev., 12, 4955–4997, https://doi.org/10.5194/gmd-12-4955-2019, 2019.

Tichý, O., Ŝmídl, V., Hofman, R., Ŝindelárová, K., Hýza, M., and Stohl, A.: Bayesian inverse modeling and source location of an unintended 131I release in Europe in the fall of 2011, Atmos. Chem. Phys., 17, 12677–12696, https://doi.org/10.5194/acp-17-12677-2017, 2017.

Tomas, J. M., Peereboom, V., Kloosterman, A., and van Dijk, A.: Detection of radioactivity of unknown origin: Protective actions based on inverse modelling, J. Environ. Radioact., 235–236, 106643, https://doi.org/10.1016/j.jenvrad.2021.106643, 2021.

Ulimoen, M. and Klein, H.: Localisation of atmospheric release of radioisotopes using inverse methods and footprints of receptors as sources, J. Hazard. Mater., 451, https://doi.org/10.1016/j.jhazmat.2023.131156, 2023.