A deep learning-driven emission estimator utilizing a mixture of experts for local wind speed situations applied to high-resolution methane imagery

Thomas Plewa[1,2], André Butz[1,3,4], Christian Frankenberg[5,6], Andrew K. Thorpe[6], Georgios Matheou[7], and Julia Marshall[2,8]

[1]Institute of Environmental Physics (IUP), Heidelberg University, Heidelberg, Germany
[2]Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[3]Heidelberg Center for the Environment (HCE), Heidelberg University, Heidelberg, Germany
[4]Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Heidelberg, Germany
[5]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA
[6]NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91125, USA
[7]School of Mechanical, Aerospace, and Manufacturing Engineering, University of Connecticut, Storrs, CT 06269, USA
[8]Leipzig Institute for Meteorology, Leipzig University, Leipzig, Germany

Correspondence: Thomas Plewa (thomas.plewa@uni-heidelberg.de)

Abstract. Methane (CH_4) is the anthropogenic greenhouse gas with the second-highest impact on the Earth's radiative budget since pre-industrial times. A substantial amount of CH_4 emissions are from the fossil fuel industry and are emitted from point-like sources that can be measured using airborne or space-based spectrometers. The precise quantification of point-source emissions has proven to be difficult, with uncertainties driven by the lack of local wind speed measurements and the task of estimating the effective wind speed of the plume. Here, we continue the development of deep learning-based methods using convolutional neural networks (CNN) to estimate emissions without the need for auxiliary wind speed information. We use a library of plumes obtained from large-eddy-simulations (LES) and realistic background noise scenes from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS-NG), used in previous studies, to generate realistic synthetic data. We suggest a mixture of experts (MoE) architecture, that is able to extract the wind speed forcing used in the LES and to estimate emission rates conditional on the wind speed present in the scenes. This allows us to integrate the concept of different wind speed scenarios into the network architecture, making the performance of the network more transparent and explainable and, while still being independent of external wind speed information, makes it possible to use external wind speed information to validate or improve emission estimates. The MoE-based network, without any external wind speed information, provides a mean absolute percentage error (MAPE) of $5.65\,\%$ for scenes with CH_4 emission rates exceeding $100\,\mathrm{kg\,h^{-1}}$, which is a $40\,\%$ improvement compared to previous implementations. The proposed network is also able to address biases at high wind speed situations, leading to almost unbiased estimates over the entire emission and wind speed domain.

1 Introduction

The two most important anthropogenic greenhouse gases (GHGs) are carbon dioxide (CO_2) and methane (CH_4) (IPCC, 2023). Due to its relatively short atmospheric lifetime, CH_4 is an attractive target for fast-acting emission reduction efforts (Ocko et al., 2021). A substantial amount of anthropogenic CH_4 emissions stems from the oil and gas sector (Saunois et al., 2020; Cusworth et al., 2022), usually in the form of emissions from point sources. Since there are, unlike for e.g. anthropogenic emissions from agriculture, clear mitigation pathways, their observation and estimation plays a key role in emission mitigation efforts.

Such observations can be performed using spaceborne or airborne imaging absorption spectroscopy. Comparisons between measurements done by area flux mappers, such as the TROPOspheric Monitoring Instrument (TROPOMI) (Veefkind et al., 2012), and inventories show an underestimation of these emissions (Alvarez et al., 2018; Lu et al., 2023; East et al., 2025). This in turn suggests that there may be additional emissions from unknown sources, such as leaks or other equipment malfunctions (Duren et al., 2019; Zavala-Araiza et al., 2017), or that the emissions included in the inventories are underestimated. Both of these possibilities require further investigation using measurements with high spatial resolution, which are able to resolve emissions on a facility scale in areas of interest, such as oil and gas basins.

Such measurements on facility level can be conducted, for large emitters, either, by multispectral or hyperspectral point source imagers such as the PRecursore IperSpettrale della Missione Applicativa (PRISMA) (Guanter et al., 2021), the Environmental Mapping and Analysis Program (EnMAP) (Roger et al., 2024), Earth Surface Mineral Dust Source Investigation (EMIT) (Thorpe et al., 2023) or spectrometers with even finer spectral resolutions such as the GHGSat constellation (Jervis et al., 2021) or planned missions such as CO2Image (Strandgren et al., 2020), with spatial resolutions on the order of tens of meters. Additionally, aircraft-based measurements with instruments such as the next-generation Airborne Visible/Infrared Imaging Spectrometer (AVIRIS-NG) (Thorpe et al., 2014, 2017), AVIRIS-3 (Green et al., 2022) or Methane Airborne MAPper 2D Light (MAMAP2D-Light) (Huhs et al., 2026) are able to detect even smaller emitters due to their spatial resolution, which goes down to a few meters.

Since point source imagers have a strongly reduced coverage due to their high spatial resolution, it is necessary to have a large amount of complementary instruments to increase the spatial and temporal coverage to be able to address the problems at hand. This demands efficient algorithms to process the large amounts of data and perform detection and emission estimation to rapidly provide actionable data to mitigate emissions.

Established methods for emission estimation of point sources include the Gaussian plume inversion (Bovensmann et al., 2010; Krings et al., 2011), the source pixel method (Jacob et al., 2016), the cross-sectional flux method (CSF) (Krings et al., 2011; Cambaliza et al., 2014; Varon et al., 2018; Kuhlmann et al., 2021; Gałkowski et al., 2025) and the integrated mass enhancement (IME) (Frankenberg et al., 2016; Varon et al., 2018). While all these methods have different areas of application due to their underlying assumptions, they all share the necessity of auxiliary wind speed information to reflect the local effective wind speed of the plume.

For emission estimation from high-resolution plume measurements, the only applicable methods are IME and CSF (Varon et al., 2018). The study also demonstrated that local measurements of the 10-meter wind speed can be used to parametrize the effective wind speed of a plume, and that using meteorological fields as input causes errors of up to 50% or 65% for IME or CSF.

Recent studies such as Eastwood et al. (2025) proposed a method involving flight maneuvers to be able to detect the same scenes multiple times. This made it possible to estimate the average wind speed acting on the plume by tracing features of the plume from one recorded scene to the next. This quantity better represents the local effective wind speed than what can usually be measured, leading to a relative deviation from the true metered emission as low as 10% with IME.

This study also illustrated the high temporal variability of the wind speed, based on local measurements from a $10\,\mathrm{m}$ sonic anemometer and a wind LIDAR, and how wind speed data from reanalysis products such as those from ERA5 and HRRR are unable to represent these fluctuations. This highlights the complexity of estimating the local effective wind speed of the plume, even if local measurements are available, and how a lack of such measurements dominates the precision and accuracy of emission estimates due to the limitations of the widely available reanalysis data.

In Jongaramrungruang et al. (2022) the application of a convolutional neural network (CNN) to emission estimation from high-resolution methane imagery was proposed and it was demonstrated that it is possible to provide accurate emission estimations without the need for external wind speed information. This idea was based on a previous study (Jongaramrungruang et al., 2019) that used IME to estimate emissions, and relied solely on the shape of the plume to estimate local wind speed properties.

Following this study, there were multiple studies exploring the application of deep learning methods, mostly CNN architectures, to emission estimation for different instruments, GHGs or specific emission scenarios: Joyce et al. (2023) for PRISMA, Radman et al. (2023) for Sentinel-2, Dumont Le Brazidec et al. (2024) for CO2M, Bruno et al. (2024) for GHGSat-C1 or Ouerghi et al. (2025) for PRISMA and EnMAP, further showcasing the potential of deep learning for emission estimation.

In our previous study (Plewa et al., 2025), we presented improvements to the method of Jongaramrungruang et al. (2022) by reducing systematic errors and adding meaningful error estimates. While the overall performance improved, our analysis revealed biases in the estimated emissions for high wind speed situations, which were also present in the previous study by Jongaramrungruang et al. (2022).

Further, we found a tendency towards clustering in the error estimates for low wind speed scenarios, which is where the network performed best, indicating that in those cases relevant information about the wind speed scenario was implicitly extracted. The stability the CNN emission estimation approach exhibited in previous studies should make it feasible to efficiently utilize widely available meteorological data.

In this study we propose a solution to these biases and make the results of the network more transparent. We aim to do so by making the assignment into different wind speed categories of the network explicit, by letting the network first estimate the wind speed conditions and then perform a conditional estimate of the emissions afterwards. This should lead to more stable and explainable estimates, still without the need for auxiliary wind speed information. In addition, external wind speed information, ideally comparable to that of ERA5 or HRRR, could be used to validate or improve emission estimates. To achieve these

features, we adopt a mixture of experts approach with independently trained components, which leads to domain-informed conditional estimates.

In Sect. 2 we provide a brief overview over the synthetic data that we used and its generation. Section 3 illustrates the changes to the network architecture and describes the training procedure that was used and its resulting properties. The performance of the neural network on the test data is discussed in Sect. 4 for applications without and also with use of external wind speed data. In the final section we provide a summary and discussion of the results of our study.

2 Data

The data used for this study were already described and presented in Jongaramrungruang et al. (2022) and Plewa et al. (2025). Since there are no changes to the underlying data, we will only provide a very brief overview and refer to the previous publications for more details.

The dataset consists of simulated plumes and background measurement from the AVIRIS-NG instrument. The plumes were created using large eddy simulations (LES), providing a temporally resolved $3-$D wind field in which a tracer, released at ground level, is transported. The LES setup is described in Matheou and Bowman (2016) and the initial parameters that were used can be found in Jongaramrungruang et al. (2019).

This dataset aims to create basic but realistic synthetic measurements of CH_4 as they could be measured by AVIRIS-NG in areas such as oil fields.

For this, the the $3-$D LES CH_4 fields are integrated vertically, weighted by the AVIRIS-NG column averaging kernel, to create column-integrated observation of the simulated tracer.

The observation conditions are inspired by conditions present during the Four-Corners campaign (Frankenberg et al., 2016): The emission rates range from $0\,\mathrm{kg\,h^{-1}}$ to $2000\,\mathrm{kg\,h^{-1}}$, the geostrophic wind speed ranges from $1\,\mathrm{m\,s^{-1}}$ to $10\,\mathrm{m\,s^{-1}}$ and the latent and sensible heat flux are $40\,\mathrm{W\,m^{-2}}$ and $400\,\mathrm{W\,m^{-2}}$.

From these LES plumes, a library consisting of 7000 CH_4 plume images, with a spatial resolution of $5 \times 5\,\mathrm{m^{-2}}$ and an overall size of $1.5 \times 1.5\,\mathrm{km^{-2}}$, was created. The plumes are equally distributed across the different wind speed scenarios. From such a library of turbulent plumes it is now possible to generate a large amount of different emission scenarios by augmenting these turbulent realizations.

Neglecting effects of self-buoyancy, the plume mass can be scaled by a random factor to simulate different emission rates of the point source. Additionally, the plumes are rotated, by between $-170°$ to $170°$, to emulate different wind speed directions and shifted by up to 30 pixels, to slightly alter the source position.

This library of augmented plumes can now be combined with a library of 3000 different retrieved scenes from AVIRIS-NG flight lines. The scenes in this library were selected to contain no point sources, but rather random and correlated noise features from urban, agricultural and desert areas and thus provide realistic background scenes for the simulated plumes. These background scenes are also rotated and shifted, and then added to the augmented plume, creating a realistic synthetic measurement.

	Number of samples		
Set	Entire dataset	Expert 1 and 10	Other experts
Training	$28000 \times \mathrm{N_{epochs}}$	$(16800 + 5600) \times \mathrm{N_{epochs}}$	$(16800 + 2 \times 5600) \times \mathrm{N_{epochs}}$
Validation	21000	$(3150 + 1050)$	$(3150 + 2 \times 1050)$
Test	17500	–	–

Table 1. Number of samples created by combining and augmenting the scenes from the plume and noise library for the training, validation and test dataset. The summation for the different experts indicates how much of the data is taken from different wind speed conditions.

As a last step, a masking threshold of 500 ppm m is applied. This value is motivated by a conservative estimate of the single-measurement precision error of AVIRIS-NG, see Jongaramrungruang et al. (2022). An example of the result of such an augmentation process can be seen in Fig. 1.

The data are split into training, validation and test data, following the 80%, 15% and 5% split from Jongaramrungruang et al. (2022) and Plewa et al. (2025). We use the same sampling strategy as described in Plewa et al. (2025), where we generate 50 scenes from each plume in the library for the test dataset and 20 for the validation set. For the training set we generate new random realizations for every epoch, with each epoch containing every scene five times. An overview of the created number of samples is depicted in Table 1.

3 Method

In this section we provide an overview of the methods we use and the changes we implemented compared to previous approaches, with respect to the neural network architecture and the training procedure. Section 3.1 covers the details regarding our neural network architecture. In Sect. 3.2 we will discuss how we trained our model and highlight how our training choices in combination with the neural network architecture leads to properties relevant for this specific application.

3.1 Network architecture

One of the first success stories of deep learning was in the field of pattern recognition, mostly using CNNs. The development of classification algorithms on datasets such as ImageNet (Russakovsky et al., 2014) led to numerous architectures achieving state-of-the-art performances in their respective era (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; Szegedy et al., 2015; He et al., 2015; Huang et al., 2016; Liu et al., 2021; Tan and Le, 2021).

The application to the regression task of emission estimation started with MethaNet (Jongaramrungruang et al., 2022), a custom-made CNN, and was followed by several different approaches (Joyce et al., 2023; Radman et al., 2023; Dumont Le Brazidec et al., 2024; Bruno et al., 2024; Ouerghi et al., 2025), utilizing different established CNN architectures, further showcasing the potential of deep learning for this task.

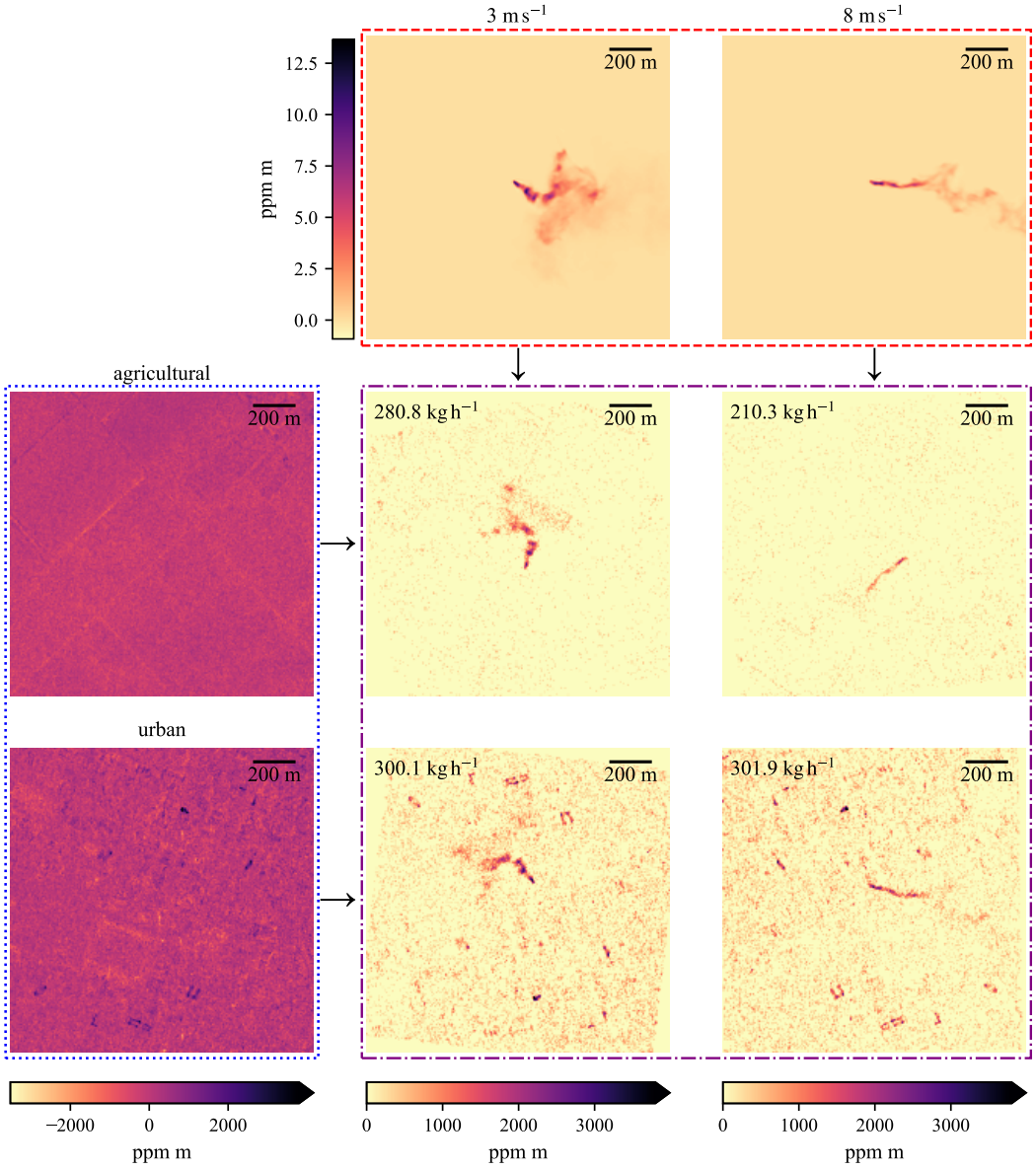


Figure 1. Example of how realistic mock scenes are created from different LES plumes and background noise scenes. In this example there are two different plumes (top) that are combined with two different background noise scenes (left) to create four different realistic scenes for analysis. The plumes are scaled, rotated and shifted. The augmented plumes are then added to the rotated and shifted background noise, and, as a last step, a threshold of 500 ppm m is applied.

Motivated by the findings in our previous study (Plewa et al., 2025), we want to tailor the network architecture further towards the specific problems of emission estimation by explicitly incorporating the concept of different wind speed scenarios into the architecture.

For this we adapt mixture of experts (MoE), an established method in machine learning (Jacobs et al., 1991). For most modern deep learning applications of MoE, the main goal is to perform conditional computations, reducing the amount of active parameters and thus the computational cost, by selecting only the best experts for each input (Fedus et al., 2022; Shazeer et al., 2017).

A MoE block consists of two main components: A router and the experts. The router selects suitable experts, conditional on the input, and the experts then process the assigned input. There are usually many of these constellations stacked and the majority of parameters of such a model are located in the experts.

Training a MoE often is a complex task and introduces a vast space of hyperparameters. However, for our application and network size we are more concerned with being able to provide conditional estimations instead of computational scaling, allowing us to substantially simplify the architecture and training.

We will only use one instance of a router followed by multiple experts, with each of these experts being a modified ResNet-50, as described in Plewa et al. (2025). The router is also a ResNet-50, following He et al. (2015).

Since we have an underlying physical theory that allows us to cluster our data by meaningful parameters, we can use our router to provide a domain-informed assignment.

In our case we let the router perform a hard assignment by classifying the wind speed forcing present in the LES. The external clustering also allows us to easily manage the amount of training data every expert sees and decouple it from the performance of the router, thus avoiding problems such as load balancing.

Therefore, this simple MoE architecture allows for independent training of each part of the entire network, which makes it easier to train and also maintains modularity for future developments. However, most importantly, it allows us to incorporate the concept of different wind speed scenarios into the network architecture. This enables the usage of external wind speed information, while still being functional in its absence.

Such information, if available and reliable, might then be used to either skip the classification step or to compare the two wind speed estimates and filter out scenes with strong deviations. An illustration of the network concept can be seen in Fig. 2.

3.2 Training setup

The MoE network that we proposed in the previous section allows for independent training of its individual components. This leads to two different training processes: One covering the training of the routing network, i.e. a classification task using the cross entropy (CE) loss function, and another one covering the training of the experts using the Gaussian negative log likelihood (GNLL) loss.

For the routing network we use the entire dataset described in Sect. 2. We trained the router using Adam (Kingma and Ba, 2017), with a batch size of 100 and an initial learning rate of 10^{-4}, for 150 epochs, followed by a second run with a learning

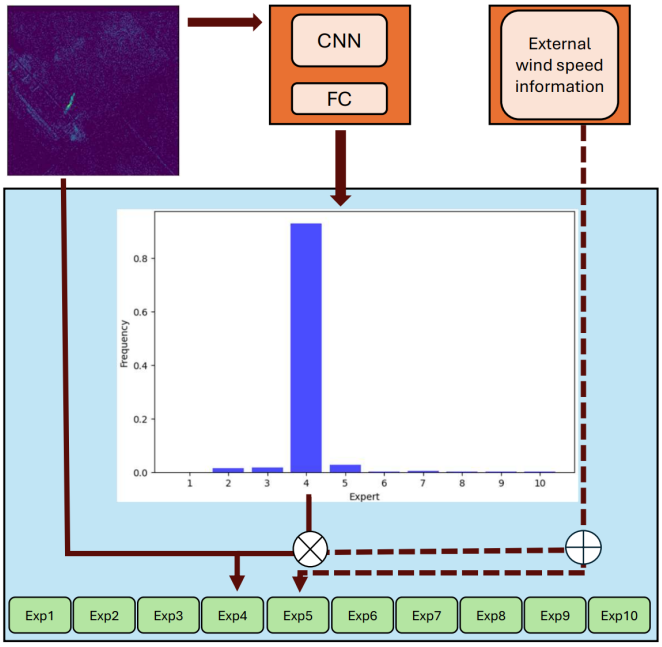


Figure 2. Illustration of the network architecture and its functionalities, showcasing the routing of the data to the different experts and how external wind speed information can be included into the process by either skipping the routing or comparing it to the estimate of the router. Each of the experts and the router in this case represent a slightly modified ResNet-50 architecture with an attached fully connected (FC) layer for the regression or classification task. The illustration shows an example where at the $\oplus$-symbol either the external wind speed information can be used to select expert 5 or the external wind speed information can be combined with the estimate of the router at the $\otimes$-symbol and used for validation/filtering.

rate of 10^{-6}. We selected the best-performing model on the validation dataset after 226 epochs, at which point the CE loss on the validation dataset started to plateau.

For the creation of the experts we only select a subset of the data, described in Sect. 2, belonging to certain wind speed scenarios.

While this allows for the specialization of an expert, it also significantly reduces the available training data. To reduce potential problems we use the already existing model, described in Plewa et al. (2025), and finetune it for the specific wind speed regime. This allows us to start from an already existing model that exhibited robustness over most of the wind speed domain.

Further, we symmetrically add data from the surrounding wind speed scenarios, using each scene from the correct wind speed forcing 30 times and each adjacent wind speed scenario 10 times. We chose the same amount of augmentations per scene for the validation data used during the training.

This leads to $40\,\%$ of the scenes being from the adjacent wind speed scenarios for all experts except for the $1\,\mathrm{m\,s^{-1}}$ and $10\,\mathrm{m\,s^{-1}}$ expert, as there is only one adjacent wind speed scenario, thus resulting in only $25\,\%$ of the scenes being from the adjacent wind speed. A summary of the training data used for each expert can be seen in Table 1.

This increases the amount of training data per expert and also makes sure that previously existing interpolation between different wind speed scenarios remains robust.

The range and the ratio of the additionally selected wind speed scenarios are selectable hyperparameters, which should make it possible to balance specialization, i.e. the performance for a specific wind speed scenario, and the robustness to different, potentially faulty, wind speed input. Therefore, each expert should represent a network that performs the emission estimation task for a local region in the wind speed domain.

In our specific case we chose only the directly adjacent wind speeds and a rather broad distribution. This reflects the assumption that available wind speed information might deviate slightly (i.e. $\pm 1\,\mathrm{m\,s^{-1}}$) and rather frequently. Such parameters should be tuneable to represent specific scenarios and different wind speed products.

We trained each expert with these data settings and a learning rate of 10^{-4} until the validation loss seemed to plateau, except for wind speeds from $6\,\mathrm{m\,s^{-1}}$ to $10\,\mathrm{m\,s^{-1}}$, where we switched to a learning rate of 10^{-6} after a few epochs. Besides these changes in learning rate and the training data distribution, we followed the training process described in Plewa et al. (2025).

Following the above-described training procedure, especially not training the model end-to-end, should have implications on the performance of the network.

While the performance of the experts are expected to be very good locally due to the isolated training, the experts' ability to provide meaningful emission estimations outside of their local wind speed domain should be limited. This is especially true for error estimates outside of the domain of an expert.

This makes the newly explicit wind speed assignment a potential bottleneck regarding the performance of the model. However, we argue that this is a desirable property of the network, as it makes external wind speed information more meaningful, allowing potentially poor estimates to be flagged, by comparing the assignment of the router to e.g. meteorological reanalysis data. It is also worth mentioning that doing end-to-end training to mitigate such effects reduced the performance of the experts as well as the overall performance, in our experience.

4 Results

The results on realistic noise scenes without the usage of any wind speed-information are presented in Sect. 4.1. In Sect. 4.2 we will analyze the results of the network and its behavior when external wind speed information is used.

Throughout these sections we will use the mean percentage error (MPE) as a summary statistic for the bias of our predictions and the mean absolute percentage error (MAPE) as a summary statistic for the spread of our data.

4.1 Application without any wind speed information

The results of the application of the neural network without any external wind speed information, i.e. relying solely on the wind speed scenario estimate of the router, can be seen in Fig. 3 a). Individual emission estimates are shown as grey points, while the mean and standard deviation of the results in $100\,\mathrm{kg\,h^{-1}}$ bins are shown in black. The scattered data and the statistics of their corresponding ensembles are tightly bound to the 1:1 line, showing a strong linear correlation between the true and estimated emission rates with only a few exceptions showing a stronger deviation.

The frequency with which the scenes with a certain wind speed forcing were assigned to different experts can be seen in Fig. 4. In addition, the fraction of the scenes with true emissions below $400\,\mathrm{kg\,h^{-1}}$ is highlighted.

The figure shows a generally high compliance between the wind speed forcing and the selected experts, especially for lower wind speed forcings. Overall, for low wind speed forcings the assignment of the wind speed class works well, and most disagreements are for lower emission scenarios. The higher the wind speed forcing becomes, the less accurate the assignments become, and the fraction of incorrect classifications not associated with low-emission scenarios increases. This indicates that high wind speed scenarios seem to be more difficult to classify correctly.

Figure 5 a) displays the relative deviations of the true emission rates from the predictions of our network. Furthermore, the accuracy of the wind speed forcing assignation for correct predictions and predictions within $\pm 1\,\mathrm{m\,s^{-1}}$ of the true value within each $100\,\mathrm{kg\,h^{-1}}$ bin are shown.

This provides a closer look at the deviations of the emission estimates over the entire emission domain and also provides us an overview over the wind speed classification performance at different emission rates. We can see that the model provides, on average, almost unbiased predictions over the entire emission range with the exception of very low emission rates.

We can also observe a slow decrease of the accuracy of the router for scenes with emission rates below $1000\,\mathrm{kg\,h^{-1}}$ that accelerates and leads to an accuracy of slightly below 0.2 for the lowest emission rates. A drop in accuracy of the router also seems to be linked to increases in the average spread of the emission estimates, thus showing a link between the router's ability to effectively categorize the wind speed scenario and the quality of the emission estimates.

It is further noteworthy that the overall performance remains rather stable until the accuracy for roughly matching wind speed conditions starts to drop, which happens slightly after the accuracy for exact classifications.

The network manages to achieve a MPE of -0.10% and a MAPE of 5.65% for scenes with an emission rate above $100\,\mathrm{kg\,h^{-1}}$, the results for emissions larger than $40\,\mathrm{kg\,h^{-1}}$ and $400\,\mathrm{kg\,h^{-1}}$ can be seen in Table 2.

Looking at the isolated performance for different wind speed forcings, as shown in Fig. 6, we can see that the model shows little to no bias over the entire wind speed forcing domain, with a tendency for increasing biases at low emission rates. The emission predictions of the model also become less precise for lower emission rates.

We can also observe a trend toward decreasing model performance for higher wind speed forcings, which leads to a very small, but consistent, bias for estimates at lower emission rates at $10\,\mathrm{m\,s^{-1}}$. This drop in performance with respect to emission estimation is also reflected in the accuracy of the wind speed classification.

While the accuracy for low wind speed situations is close to 1.0 and stays rather stable for scenes till around $400\,\mathrm{kg\,h^{-1}}$, it becomes unstable around $600\,\mathrm{kg\,h^{-1}}$ for higher wind speed scenarios.

For $8\,\mathrm{m\,s^{-1}}$ to $10\,\mathrm{m\,s^{-1}}$ we can see that the classification performance is notably worse even for higher emission rates. For $9\,\mathrm{m\,s^{-1}}$ to $10\,\mathrm{m\,s^{-1}}$, the accuracy for roughly matching wind speed classifications at low emission rates increases and for $10\,\mathrm{m\,s^{-1}}$, even the accuracy rate for the correct classification improves.

This could be explained by the fact that for high wind speed situations the emitted mass is transported more quickly out of the scene. This results in more scenes at low emission rates with barely any trace of a signal above the applied concentration threshold left for high wind speed situations. Since there is barely any signal left with which to perform an emission estimation, the network essentially has to guess the wind speed situation for the most part and since high wind speed scenarios are the most likely, they are the favorable estimate.

This inflates accuracies for correct classifications for $10\,\mathrm{m\,s^{-1}}$ and also for almost correct classifications for $9\,\mathrm{m\,s^{-1}}$. Signs of this behavior are also visible in Fig. 4 by the increased assignment of low emission scenes to the $10\,\mathrm{m\,s^{-1}}$-expert.

Figure 7 a) shows the distribution of the difference between our estimated emissions and the true emissions divided by their corresponding estimated standard deviation. Thus, the data should ideally follow a normal distribution with mean of 0.0 and a standard deviation of 1.0, if all the error estimates are meaningful.

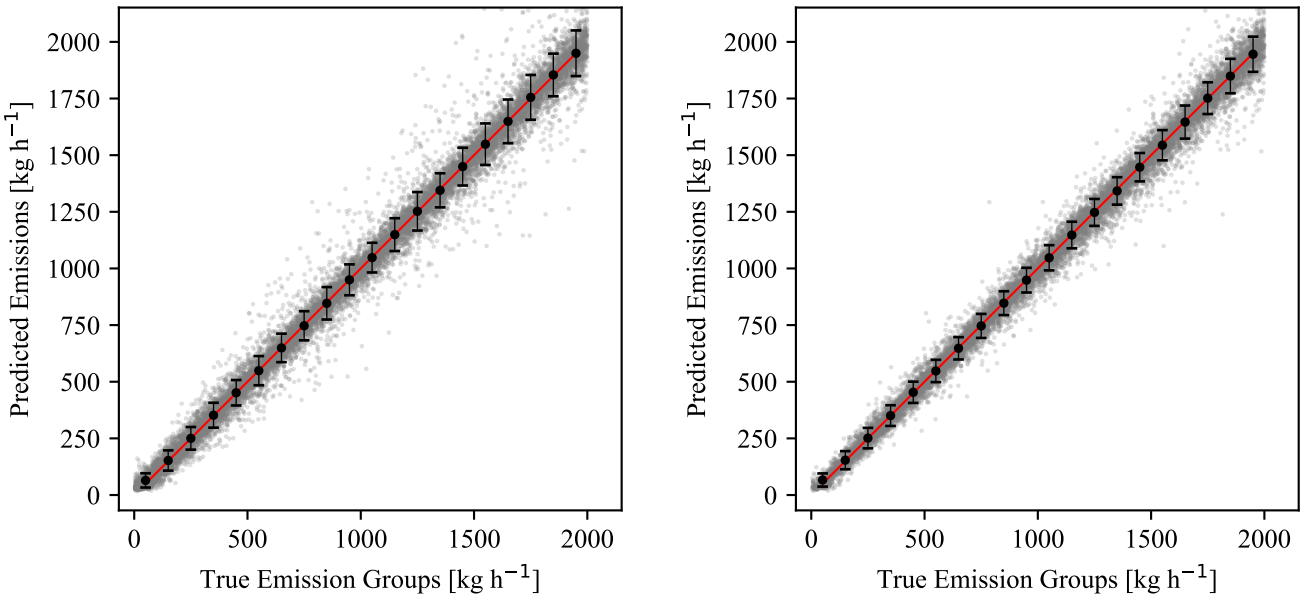
Following the distribution shown by the histogram, we can see that the data follow a normal distribution, with the exception of a few outliers. These outliers, however, cause the best fit to look far from optimal. Given our network design and the training procedure, described in Sect. 3, the error estimates for scenes that deviate by multiple $\mathrm{m\,s^{-1}}$ from the true wind speed situation will most likely not be meaningful, which makes such outliers an expected consequence.

4.2 Application using wind speed information

In this section we make use of the properties of our network architecture and training process, described in Sect. 3, which makes the wind speed assignment a critical point for the emission estimation performance, by combining external wind speed information with the classification results from the router. To this end, we filter out results with a deviation larger than $1\,\mathrm{m\,s^{-1}}$ from the true underlying wind speed forcing, thus by assuming that we have spatially averaged, roughly correct external wind speed information.

The filtered version of the emission estimates against the underlying true emission rate along with the means and standard deviations of their corresponding emission ensembles are depicted in Fig. 3 b). Compared to the unfiltered version, we can see that most of the outliers are effectively filtered out over the entire emission domain, however, the effect is especially visible for low emission rates.

Looking at Fig. 5 b) we can see that, as an effect of filtering, the accuracy for roughly matching wind speed conditions is forced to 1.0 and the frequency of correct classifications remains stable for emissions down to around $800\,\mathrm{kg\,h^{-1}}$. For emissions smaller than $800\,\mathrm{kg\,h^{-1}}$ the accuracy slowly starts to decay and the rate of decay accelerates from there, leading to an accuracy of around 0.4 for scenes with the lowest emission rate.



(a) Unfiltered emission estimates

(b) Filtered emission estimates

Figure 3. Comparison of the predicted emissions against their ground truth, showing the scatter of the estimates (grey dots) and, for a more quantitative view, the mean and standard deviation of emission ensembles (black dots and error bars). The ensembles have been created by separating the data into groups containing scenes in $100\,\mathrm{kg\,h^{-1}}$ bins. The left plot shows the unfiltered emission estimates, not assuming any external wind speed information, and the right plot uses external wind speed information to filter out scenes with a strong deviation in their respective wind speed assignment. The $1:1$ line is shown in red.

We can also see that the predictions of the model over the entire emission domain show little to no bias and a steady development of the ensemble spread, leading to an overall more stable performance. This is also represented by its MPE of 0.16% and MAPE of 4.26% for scenes larger than $100\,\mathrm{kg\,h^{-1}}$, the values for scenes larger than $40\,\mathrm{kg\,h^{-1}}$ and $400\,\mathrm{kg\,h^{-1}}$ are included in Table 2, and also show a more stable behavior of the change of the MPE.

The relative deviations from the predicted emissions to the true emissions for the different wind speed ensembles are shown in Fig. 8. Overall the inclusion of wind speed information shows a positive effect, reducing biases and the spread of the data, especially regarding the spread of the emission ensembles for low emission rates.

The slight tendency towards underestimations of the expert responsible for wind speed forcings of $10\,\mathrm{m\,s^{-1}}$ is reduced even further. However, for $7\,\mathrm{m\,s^{-1}}$, $8\,\mathrm{m\,s^{-1}}$ and $9\,\mathrm{m\,s^{-1}}$ we can observe a slight trend towards overestimation of emissions in the $100\,\mathrm{kg\,h^{-1}}$ to $200\,\mathrm{kg\,h^{-1}}$ range. Besides that, the estimates show little to no bias over the entire emission domain

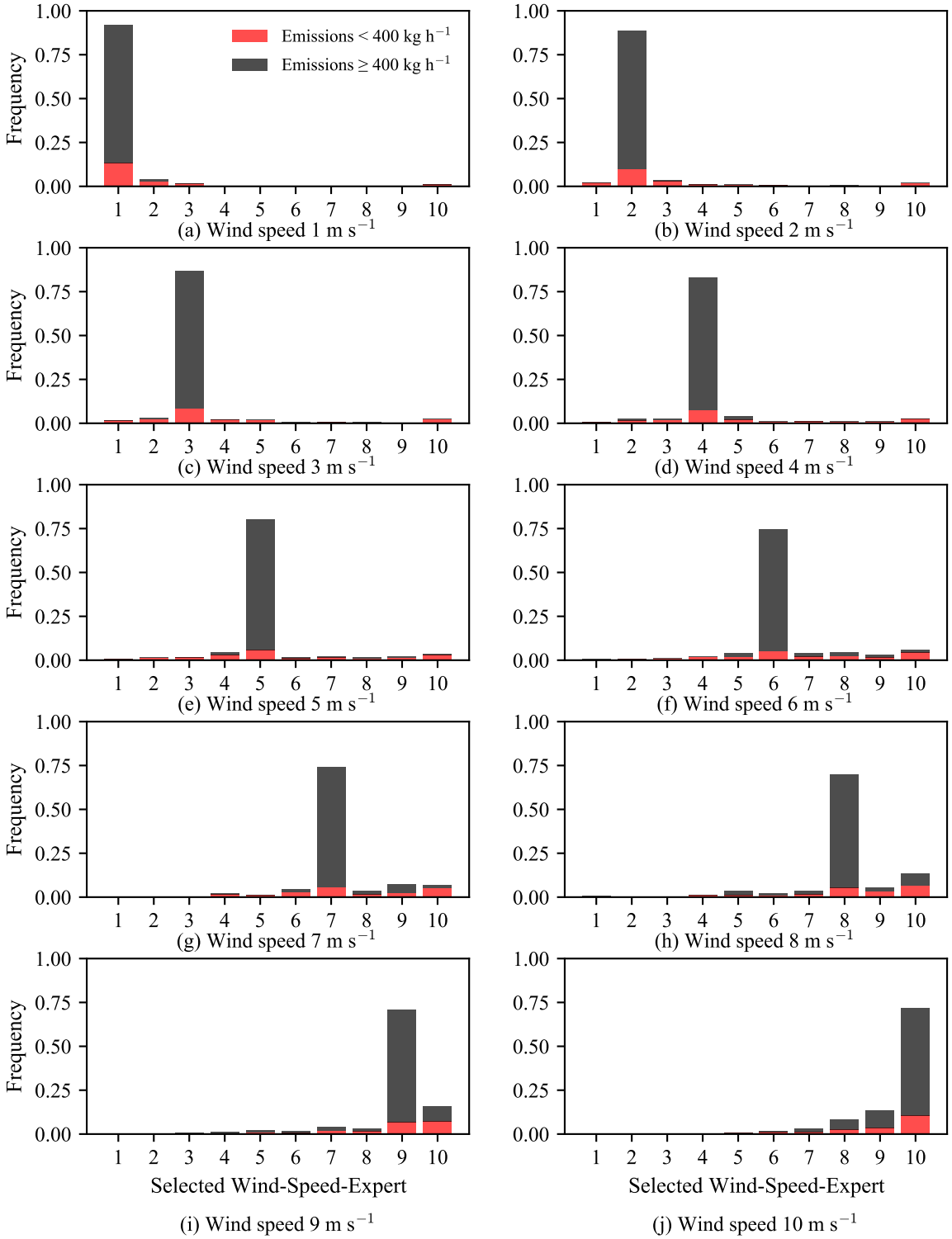
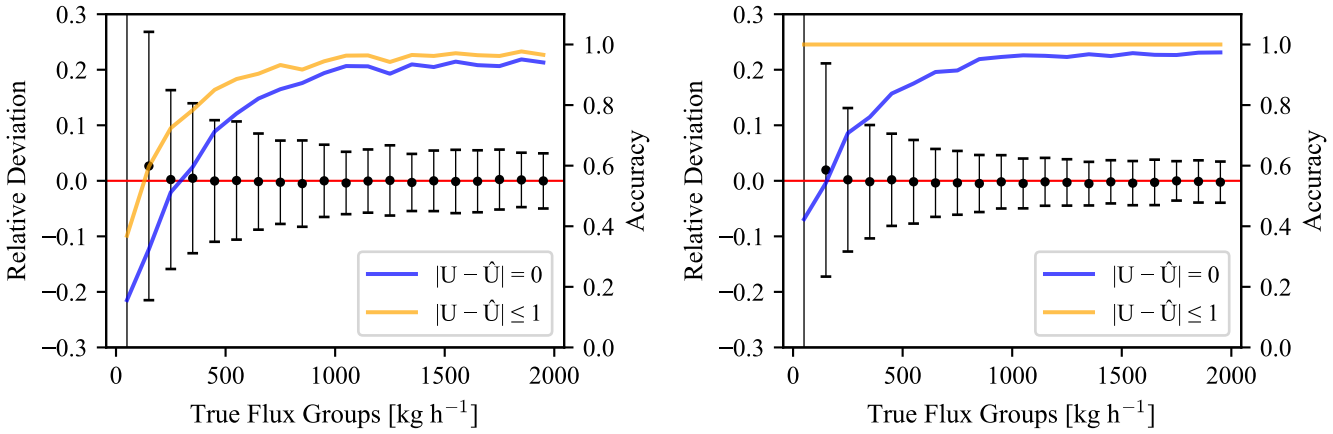


Figure 4. Frequencies with which scenes from different wind speed situations get assigned to the different experts. The scenes with emissions below $400\,\mathrm{kg\,h^{-1}}$ are displayed in red.



a) Unfiltered relative deviations

(b) Filtered relative deviations

Figure 5. Relative deviations of the predicted emission rates, separated into $100\,\mathrm{kg\,h^{-1}}$ bins, against their ground truth. The mean and standard deviation of the relative deviations are shown in black. Additionally, the accuracy (right axis) of the router for correct wind speed assignments (in blue) and those that are within $\pm 1\,\mathrm{m\,s^{-1}}$ of the truth (in orange) are displayed. The left plot shows the unfiltered predictions, assuming no external wind speed information, and the right plot shows the effect of filtering for wind speed estimates that are within $\pm 1\,\mathrm{m\,s^{-1}}$ of the true wind speed.

even for isolated wind speed forcings, including the expert for $10\,\mathrm{m\,s^{-1}}$ wind speed forcing. The accuracy of the wind speed assignments still shows a correlation with the quality of the emission estimates.

The deviations of the filtered estimated emissions from the true emissions with respect to their corresponding error estimates, displayed in Fig. 7 b), follow a normal distribution, indicating meaningful error estimates for roughly matching wind speed scenarios. Even for isolated wind speed scenarios the error estimates look meaningful for almost all wind speed forcings, see for example Figs. 9 a), 9 b) and 9 c). The expert responsible for scenes with a wind speed forcings of $8\,\mathrm{m\,s^{-1}}$ shown in Fig. 9 b) shows the largest deviation with a few cases that show underestimated errors, however, the overall distribution still looks good.

So far we have only filtered out the results with large deviations in their wind speed classifications and showed that this improves the performance.

This, however, does not imply that the performance of the network for those scenes would improve when the wind speed classification is corrected by using external wind speed information.

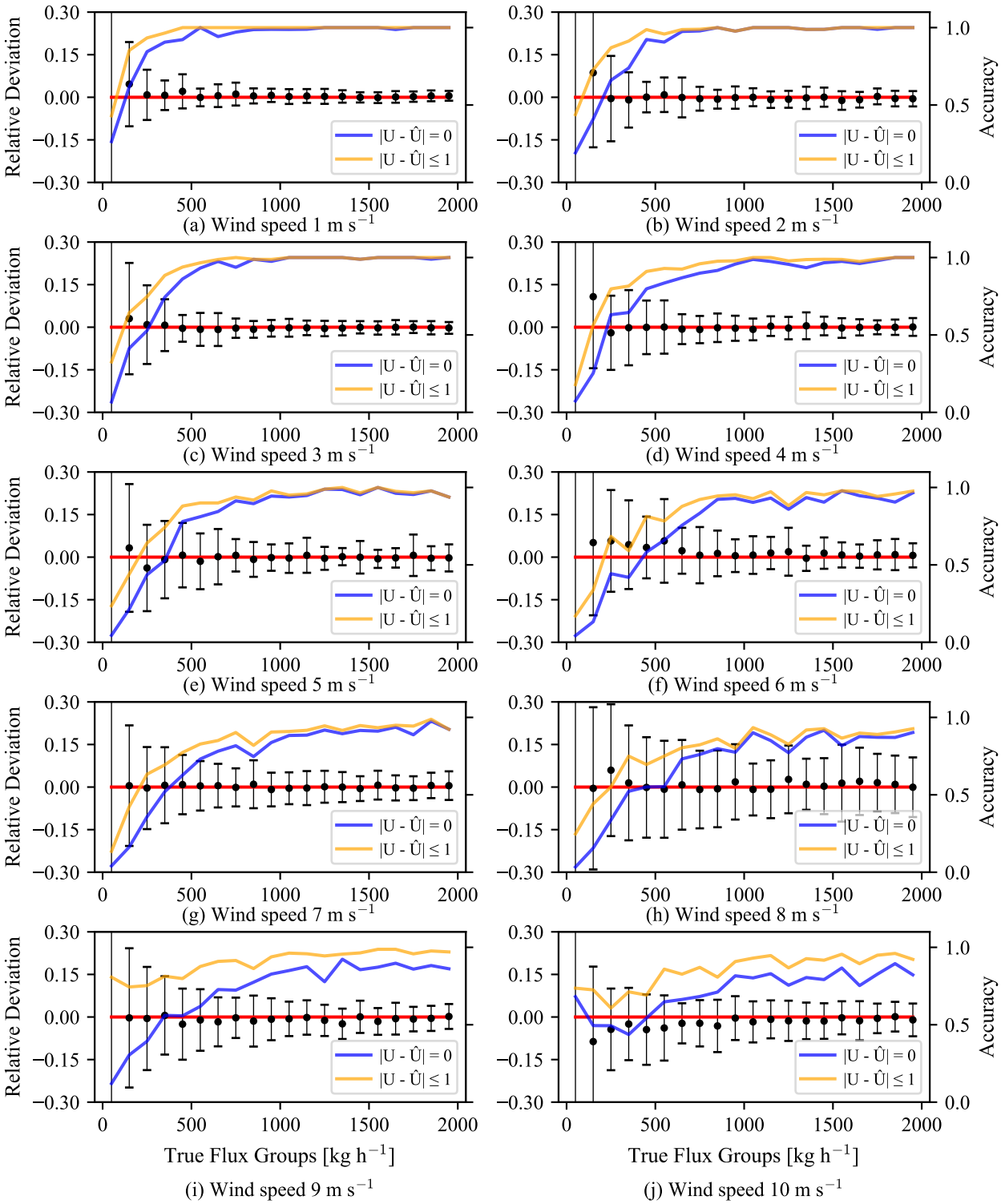
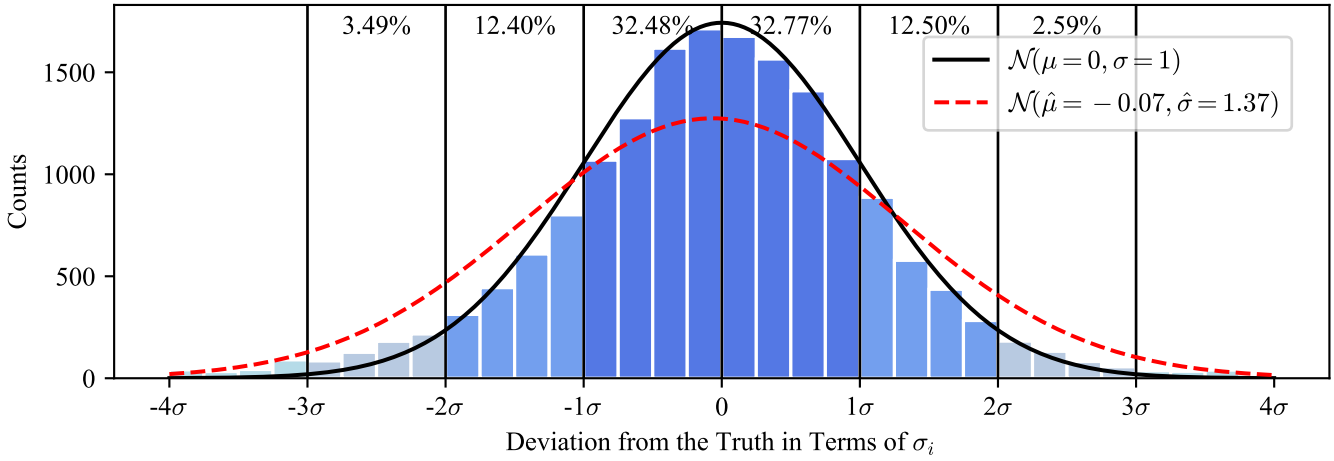
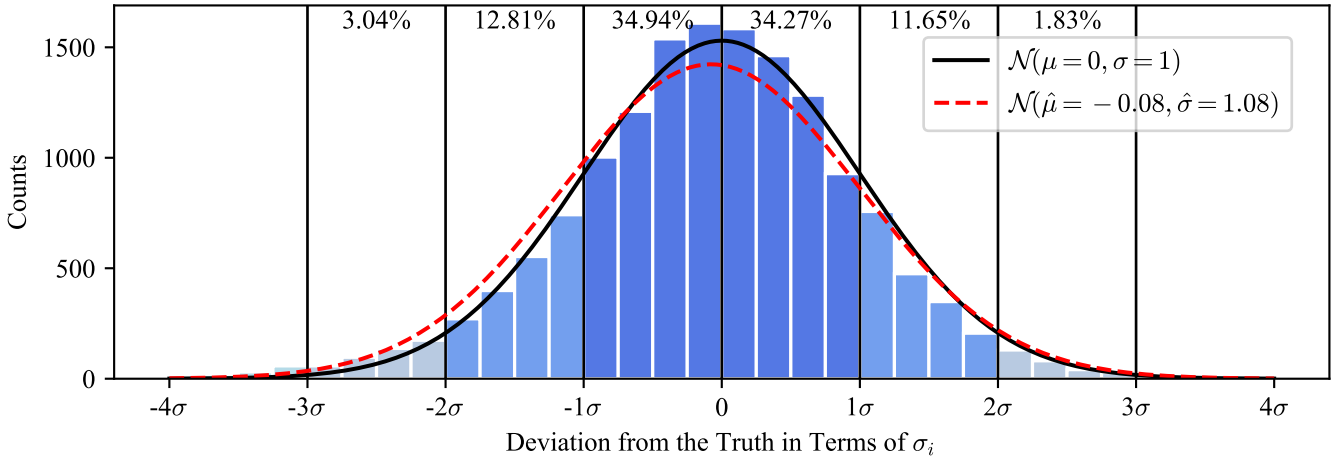


Figure 6. Relative deviations of emission rate ensembles, spanning a range of 100 kg h^{-1}, of the estimated emission rates from the true emission rates for different wind speed conditions. In addition, the accuracy of the wind speed assignments for correct and ±1 m s^{-1} correct estimates is displayed.



(a) Unfiltered estimates



(b) Filtered estimates

Figure 7. Deviations of the predicted from the true emission rates with respect to their corresponding uncertainty estimate for each scene for an unfiltered case (upper panel) and for scenes filtered for roughly matching wind speed scenarios (lower panel). The ideal distribution is depicted in black and the best fit to the data in red.

We can investigate this by assigning the scene to the expert suggested by the external wind speed information instead of the estimated one. This can be done for all scenes or for scenes that show absolute deviations larger than $1\,\mathrm{m\,s^{-1}}$, further it can be done using the correct wind speed forcings or ones that show deviations from the correct one.

In this case we assume that external wind speed information will be correct $\pm\,1\,\mathrm{m\,s^{-1}}$. By applying the slightly wrong wind speed information to the network we are also able to probe the stability of our expert with respect to the accuracy of the external input.

Figure 10 a) shows the behavior of the MPE and MAPE of the network for scenes with an emission rate larger than a certain threshold with all scenes directed to either the correct expert or the expert responsible for scenes with a wind speed forcing larger or smaller by $1\,\mathrm{m\,s^{-1}}$.

This provides an overview over the stability of the predictions and shows that the impact of incorrect external wind speed information is stronger for smaller emission rates and then gradually becomes weaker, showing that the MPE for the overall performance quickly falls below an absolute deviation of around 2% and then gradually decreases to around 1% for overestimated wind speed forcings, and even lower for underestimated wind speed forcings. The MAPE behaves similarly and steadily falls from around 7% to around 5%.

While it is an extreme assumption to only have over- or underestimated wind speed forcings, it showcases the stability of the experts and demonstrates that the training methods described in 3.2 lead to a stable expert. For a more realistic scenario, we chose to only replace the scenes that showed a notable deviation to begin with. The results can be seen in Fig. 10 b). We can observe that the only visible impact is for small emission scenarios, and even there the deviations are barely notable.

For a more quantitative comparison, and to show that the quality of the estimates could be improved by using external wind speed information, we can look at the MPE and MAPE for all scenes above $40\,\mathrm{kg\,h^{-1}}$, $100\,\mathrm{kg\,h^{-1}}$ and $400\,\mathrm{kg\,h^{-1}}$, shown in Table 2.

We can compare the performance of the network only using the correct wind speed classification, the network using the correct ones only for scenes that showed a notable deviation, the filtered version, and the unfiltered version. From here on and in Table 2 these are referred to as "correct", "corrected", "filtered", and "unfiltered", respectively.

This comparison reveals that the version with filtered results shows the best performance. The version using the correct wind speed classes performs slightly better than the corrected version, while all outperform the unfiltered version.

This behavior clearly shows that using external wind speed information to correct the wind speed classification can increase the quality of the emission estimates, even for scenes which were not initially assigned correctly, and thus shows that the design properties of the network transition well into application. The fact that the filtered version shows the best performance indicates that scenes which are difficult to classify in terms of their wind speed situation are also, in addition, difficult for accurate emission estimation, even if their wind speed class is known.

However, it is worth noting that this effect is exaggerated when low emission rates are included, since they are overall more difficult to process. Thus, removing them increases the overall performance. For scenes larger than $400\,\mathrm{kg\,h^{-1}}$ the performance of the filtered version is already almost on par with the correct or corrected version.

As a last check regarding the performance of the network, we apply it to three controlled release experiments that have previously been used in Jongaramrungruang et al. (2019) and Jongaramrungruang et al. (2022). The scenes and their corresponding emission rates ($(39 \pm 5)\,\mathrm{kg\,h^{-1}}$, $(39 \pm 5)\,\mathrm{kg\,h^{-1}}$ and $(89 \pm 4)\,\mathrm{kg\,h^{-1}}$) as well as the wind speed measurements provided in Jongaramrungruang et al. (2019, 2022) are shown in Fig. 11. We down-scaled the scenes from their original resolution to our

expected resolution of $5 \times 5\,\mathrm{m}^{-2}$ by area averaging the measured enhancements and we apply the 500 ppm m threshold to make them comparable to the training data. Since all three scenes are at low emission rates, the scope of this test is limited, as is the expected performance of the network. The network assigns wind speed classes of $4\,\mathrm{m\,s}^{-1}$, $3\,\mathrm{m\,s}^{-1}$ and $4\,\mathrm{m\,s}^{-1}$ alongside emission estimates of $(82.9 \pm 28.6)\,\mathrm{kg\,h}^{-1}$, $(104.6 \pm 24.1)\,\mathrm{kg\,h}^{-1}$ and $(123.8 \pm 28.3)\,\mathrm{kg\,h}^{-1}$, respectively. The accuracy of the wind speed classification at this emission rate is quite low and thus no longer trustworthy, therefore we are relying on the external wind speed measurements performed at the site of the controlled release: $0.9\,\mathrm{m\,s}^{-1}$, $0.8\,\mathrm{m\,s}^{-1}$ and $1.6\,\mathrm{m\,s}^{-1}$. Thus, for the first two scenes we use the $1\,\mathrm{m\,s}^{-1}$ expert and for the last scene we use a mixture of the $1\,\mathrm{m\,s}^{-1}$ and the $2\,\mathrm{m\,s}^{-1}$ expert by linearly interpolating between the two results. This leads to emission estimates of $(54.1 \pm 10.8)\,\mathrm{kg\,h}^{-1}$, $(56.8 \pm 13.2)\,\mathrm{kg\,h}^{-1}$ and $(116.0 \pm 19.7)\,\mathrm{kg\,h}^{-1}$, respectively. The estimates are as expected based on the performance on synthetic data (see Fig. 8), in that they are slightly overestimated. If one would assume an overestimation of around $10\,\%$, the estimates are either close to or within the 1σ range of the true emission. Therefore, the network is able to provide suitable emission estimates when using the external wind speed data even for such low emission rates. As a comparison, the emission estimate using IME (Source: data.carbonmapper.org) are $(193 \pm 59)\,\mathrm{kg\,h}^{-1}$, $(126 \pm 75)\,\mathrm{kg\,h}^{-1}$ and $(108 \pm 47)\,\mathrm{kg\,h}^{-1}$, respectively. The scenes can be found under the plume names ang20171011t191143-A, ang20171011t200410-A and ang20171011t203837-A in the carbon mapper database.

5 Summary & Conclusions

In this study, we propose a novel deep learning approach for emission estimation from point sources and apply it to synthetic high-resolution imagery of methane plumes. We present a flexible mixture of experts architecture that allows the concept of different wind speed scenarios to be incorporated into the network architecture.

The network still does not require external wind speed information, but can use it if available, as it provides insight into the assumed wind speed that was used for the emission estimation.

The synthetic data were created by combining LES plumes and realistic background noise measurements, taken from flights with the AVIRIS-NG instrument, to create realistic mock measurements. The LES were created using wind speed forcings ranging from $1\,\mathrm{m\,s}^{-1}$ to $10\,\mathrm{m\,s}^{-1}$ and we aimed at estimating emission from $0\,\mathrm{kg\,h}^{-1}$ to $2000\,\mathrm{kg\,h}^{-1}$.

The network was trained to estimate the wind speed forcing present in the scene and, conditional on that wind speed forcing, to predict an emission estimate alongside an uncertainty estimate. We present a procedure that allows the mixture of experts network to be trained in such a way that the wind speed forcing should be usable as a quality assessment parameter for the predicted emission estimates. Further, the training procedure should allow the stability of the emission estimates to be balanced with respect to the wind speed assignment by tuning the degree of specialization on certain wind speed scenarios.

Applied to the test dataset, the network showed a mean percentage error (MPE) of -0.44% and -0.10% for scenes with an emission rate larger than $40\,\mathrm{kg\,h}^{-1}$ or $100\,\mathrm{kg\,h}^{-1}$ and a mean absolute percentage error (MAPE) of 6.53% and 5.65%, respectively. These results were achieved without the use of any external wind speed information and thus are comparable to the results we achieved in Plewa et al. (2025), showing a relative improvement of $40\,\%$ for scenes above $100\,\mathrm{kg\,h}^{-1}$. The

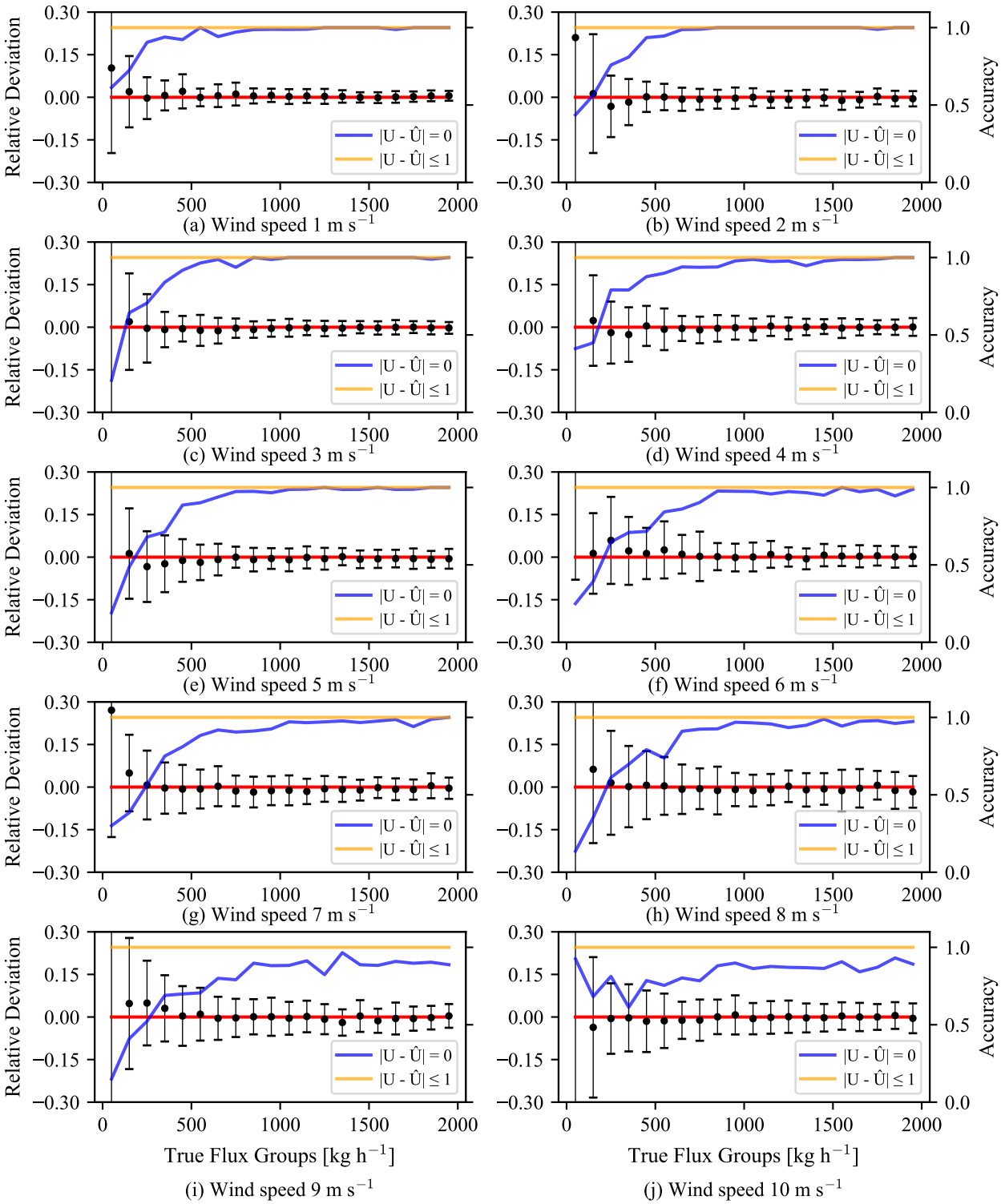


Figure 8. Relative deviations of emission rate ensembles, spanning a range of $100\,\mathrm{kg\,h^{-1}}$, of the estimated emission rates from the true emission rates for different wind speed conditions. In addition, the accuracy of the wind speed assignments for correct and $\pm 1\,\mathrm{m\,s^{-1}}$ correct estimates is displayed. The data are filtered for wind speed assignments that are within $\pm 1\,\mathrm{m\,s^{-1}}$ of the true wind speed.

predictions of the network also show even less bias than the previous version, which can mostly be attributed to an increased stability over almost the entire emission domain, with the exception of very low emission rates, which are unavoidable, see Plewa et al. (2025).

The router is able to correctly assign the majority of the scenes to their corresponding wind speed forcings. However, the performance depends on the emission rate and the wind speed forcing. The accuracy is high for high emission rates but decreases for lower emission rates, and also overall for high wind speed forcings. It shows a nice correlation to the performance of the emission rate estimation of this network, but also matches the performances of our previous network, where underlying wind speed assumptions were either not accessible or nonexistent, providing a potential explanation for its behavior.

The performance of the network for the different wind speed ensembles shows that the predictions are also almost unbiased over the entire wind speed domain, with only slight instabilities for $10\,\mathrm{m\,s^{-1}}$.

The reason for this drop in performance for high wind speed situations is not yet fully determined. Following Ouerghi et al. (2025) there should be an impact caused by overfitting of the network due to an increased correlation between scenes at low wind speed scenarios, that would decrease with increasing wind speeds. This would suggest that for future training datasets the sampling time should be scaled with the wind speed forcing to mitigate this effect. However, the rising difficulties in emission estimation, even with a known wind speed scenario, might still be related to a lack of distinguishable patterns, causing an increased uncertainty on the local effective wind speed estimate. Further, it could also be affected by the spatial limitation of the scene, since for higher wind speeds the plume mass gets transported faster out of the scene. Therefore, parts of the plume with less influence of initial instabilities of the local wind speed at the time of emission, which can play a crucial role for emission estimation, might no longer be above the detection threshold or even no longer inside the scene. Such effects were observed in the study of Gałkowski et al. (2025) for the cross-sectional flux method and might also be a factor in deep learning driven approaches.

However, overall, our network showed little to no bias over the entire emission and wind speed domain and clear improvements in both MPE and MAPE compared to the previously published version (Plewa et al., 2025), showing how effective the MoE approach is. The network also provides meaningful error estimates with the exception of a few outliers, which, given its design, is to be expected. These outliers, however, make the average uncertainty predictions over the entire dataset of the previous network better calibrated as it is closer to the expected standard normal distribution with an estimated mean and standard deviation of -0.14 and 1.02, respectively. However, given the overall increased performance of the MoE network, the distributions of the uncertainty estimates are a lot sharper.

In addition, we demonstrated that the MoE approach allows external wind speed information to be utilized. Within this study we used the geostrophic wind speed forcing applied in the LES, thus a coarse spatial average, and we assumed that the external information is $\pm\,1\,\mathrm{m\,s^{-1}}$ accurate, which is reflected in the training for our experts.

We showed that by using external wind speed information we can reduce the MAPE by between 1.07% and 1.88% for scenes larger than $40\,\mathrm{kg\,h^{-1}}$, depending on whether we correct the wind speed assignment for scenes showing a deviation larger than $1\,\mathrm{m\,s^{-1}}$ or just filter them. We also demonstrated that using external wind speed information to filter out scenes with potentially poor performance allowed further stabilization of the performance of the model over the emission and wind speed domain. This

filtering also effectively removed outliers in the error estimation, leading to even better-fitting error distributions, providing meaningful error estimates down to a wind-speed-ensemble level. Thus, by filtering the quality of the overall calibration of the error estimates becomes comparable to the previous end-to-end trained network, while improving the performance on a wind speed ensemble level especially for high wind speed situations, such as the $9\,\mathrm{m\,s^{-1}}$ case, where the MoE provides better suited error estimates, with an estimated mean and standard deviation of -0.03 and 1.03 compared to -0.42 and 1.13 in the previous version.

By short-cutting the router and directly selecting the expert, either for all scenes or only the ones with a strong deviation to begin with, we probed the stability of the experts with respect to auxiliary wind speed information. This illustrated that, even for the extreme case, with all scenes having an over- or underestimated wind speed forcing, the predictions remained rather stable with a MPE and MAPE of below $2\,\%$ and $7\,\%$, respectively, for scenes above $20\,\mathrm{kg\,h^{-1}}$.

For more realistic cases, where we only replaced the scenes with absolute deviations larger than $1\,\mathrm{m\,s^{-1}}$, there was barely a notable effect, even for very low emission rates.

Therefore, we were able to show that the properties imposed by our network architecture and training procedure transition well into application. Allowing for a network that provides a strict improvement to a previous comparable version, providing new state-of-the-art emission estimations on comparable synthetic data without the need for external wind speed information, while increasing the explainability of the network performance by revealing the wind speed assignment used for the emission estimation.

It effectively utilizes the stability of neural networks for emission estimation and makes it possible to combine it with coarse spatial wind speed information to further improve its performance, or to simply validate its results with an external source. In future applications, it should be possible to extract information from the LES that is more representative of available meteorological fields such as e.g. ERA5 or HRRR. Using this information during training would allow for a seamless usage of such data.

In a small first demonstration on three controlled release experiments, we could also show that the general network design seems to transition well into the application on real data and that the network can provide well matching emission estimates in low emission scenarios when using external wind speed measurements. Following the data-driven nature of neural networks and since the architecture is conceptually close enough to other successful applications, we believe that these improvements should be applicable within a wide range of different settings. This includes applications to satellite missions, given the comparable results presented in e.g. Ouerghi et al. (2025). While future goals for this work should include improving limitations of the current setup, such as the error estimates outside of the expert domain by e.g. using an end-to-end trained expert solely for error estimates or using finetuned ConvNext (Liu et al., 2022) architectures as a base network to improve the router as well as the experts, we would argue that the current setup provides most of the desirable qualities for emission estimations. Thus, the priority for future works would be to focus on studying the applications to real data and potential problems arising due to the more complicated environment and use these challenges to drive future conceptual changes.

	unfiltered		filtered		corrected		correct	
threshold	MPE	MAPE	MPE	MAPE	MPE	MAPE	MPE	MAPE
$\mathrm{kg\,h^{-1}}$	%	%	%	%	%	%	%	%
40	−0.44	6.53	0.09	4.65	0.13	5.46	0.19	5.31
100	−0.10	5.65	0.16	4.26	0.11	4.74	0.15	4.62
400	0.09	4.21	0.26	3.43	0.19	3.62	0.21	3.53

Table 2. Mean percentage error and mean absolute percentage error for all scenes with an emission rate larger or equal to the given threshold, using no external wind speed information (unfiltered), using external wind speed information to either filter or correct absolute deviations larger than $1\,\mathrm{m\,s^{-1}}$ (filtered or corrected, respectively) or using the correct wind speed assignment for all scenes (correct).

Data availability. The AVIRIS-NG data is from Carbon Mapper (2017) https://data.carbonmapper.org [last access: 25 August 2026]. The used LES are publicly available at https://doi.org/10.5281/zenodo.22775616

Author contributions. Thomas Plewa: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. André Butz: Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization. Christian Frankenberg: Writing – review & editing, Resources, Conceptualization. Andrew K. Thorpe: Writing – review & editing, Data curation. Georgios Matheou: Writing - review & editing, Data curation. Julia Marshall: Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

Competing interests. Some authors are members of the editorial board of *Atmospheric Measurement Techniques*.

Acknowledgements. This work was supported by the BMWK-funded project CO2KI (FZK50EE2212). This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project ID bd1231 and bb1170. The authors gratefully acknowledge the SDS@hd data storage service, supported by the Ministry of Science, Research and the Arts Baden-Württemberg (MWK).

References

Alvarez, R. A., Zavala-Araiza, D., Lyon, D. R., Allen, D. T., Barkley, Z. R., Brandt, A. R., Davis, K. J., Herndon, S. C., Jacob, D. J., Karion, A., et al.: Assessment of methane emissions from the US oil and gas supply chain, Science, 361, 186–188, 2018.

Bovensmann, H., Buchwitz, M., Burrows, J. P., Reuter, M., Krings, T., Gerilowski, K., Schneising, O., Heymann, J., Tretner, A., and Erzinger, J.: A remote sensing technique for global monitoring of power plant CO_2 emissions from space and related applications, Atmospheric Measurement Techniques, 3, 781–811, https://doi.org/10.5194/amt-3-781-2010, 2010.

Bruno, J. H., Jervis, D., Varon, D. J., and Jacob, D. J.: U-Plume: automated algorithm for plume detection and source quantification by satellite point-source imagers, Atmospheric Measurement Techniques, 17, 2625–2636, https://doi.org/10.5194/amt-17-2625-2024, 2024.

Cambaliza, M. O. L., Shepson, P. B., Caulton, D. R., Stirm, B., Samarov, D., Gurney, K. R., Turnbull, J., Davis, K. J., Possolo, A., Karion, A., Sweeney, C., Moser, B., Hendricks, A., Lauvaux, T., Mays, K., Whetstone, J., Huang, J., Razlivanov, I., Miles, N. L., and Richardson, S. J.: Assessment of uncertainties of an aircraft-based mass balance approach for quantifying urban greenhouse gas emissions, Atmospheric Chemistry and Physics, 14, 9029–9050, https://doi.org/10.5194/acp-14-9029-2014, 2014.

Cusworth, D. H., Thorpe, A. K., Ayasse, A. K., Stepp, D., Heckler, J., Asner, G. P., Miller, C. E., Yadav, V., Chapman, J. W., Eastwood, M. L., et al.: Strong methane point sources contribute a disproportionate fraction of total emissions across multiple basins in the United States, Proceedings of the National Academy of Sciences, 119, e2202338 119, 2022.

Dumont Le Brazidec, J., Vanderbecken, P., Farchi, A., Broquet, G., Kuhlmann, G., and Bocquet, M.: Deep learning applied to CO_2 power plant emissions quantification using simulated satellite images, Geoscientific Model Development, 17, 1995–2014, https://doi.org/10.5194/gmd-17-1995-2024, 2024.

Duren, R. M., Thorpe, A. K., Foster, K. T., Rafiq, T., Hopkins, F. M., Yadav, V., Bue, B. D., Thompson, D. R., Conley, S., Colombi, N. K., et al.: California's methane super-emitters, Nature, 575, 180–184, 2019.

East, J. D., Jacob, D. J., Jervis, D., Balasus, N., Estrada, L. A., Hancock, S. E., Sulprizio, M. P., Thomas, J., Wang, X., Chen, Z., et al.: Worldwide inference of national methane emissions by inversion of satellite observations with UNFCCC prior estimates, Nature Communications, 16, 11 004, 2025.

Eastwood, M. L., Thompson, D. R., Green, R. O., Fahlen, J. E., Adams, T. J., Brandt, A. R., Brodrick, P. G., Chlus, A., Kort, E. A., Reuland, F., and Thorpe, A. K.: Direct measurement of plume velocity to characterize point source emissions, Proceedings of the National Academy of Sciences, 122, e2507350 122, https://doi.org/10.1073/pnas.2507350122, 2025.

Fedus, W., Zoph, B., and Shazeer, N.: Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity, Journal of Machine Learning Research, 23, 1–39, 2022.

Frankenberg, C., Thorpe, A. K., Thompson, D. R., Hulley, G., Kort, E. A., Vance, N., Borchardt, J., Krings, T., Gerilowski, K., Sweeney, C., et al.: Airborne methane remote measurements reveal heavy-tail flux distribution in Four Corners region, Proceedings of the national academy of sciences, 113, 9734–9739, 2016.

Gałkowski, M., Marshall, J., Fuentes Andrade, B., and Gerbig, C.: Impact of atmospheric turbulence on the accuracy of point source emission estimates using satellite imagery, Atmospheric Chemistry and Physics, 25, 13 831–13 848, https://doi.org/10.5194/acp-25-13831-2025, 2025.

Green, R. O., Schaepman, M. E., Mouroulis, P., Geier, S., Shaw, L., Hueini, A., Bernas, M., McKinley, I., Smith, C., Wehbe, R., Eastwood, M., Vinckier, Q., Liggett, E., Zandbergen, S., Thompson, D., Sullivan, P., Sarture, C., Van Gorp, B., and Helm-

linger, M.: Airborne Visible/Infrared Imaging Spectrometer 3 (AVIRIS-3), in: 2022 IEEE Aerospace Conference (AERO), pp. 1–10, https://doi.org/10.1109/AERO53065.2022.9843565, 2022.

Guanter, L., Irakulis-Loitxate, I., Gorroño, J., Sánchez-García, E., Cusworth, D. H., Varon, D. J., Cogliati, S., and Colombo, R.: Mapping methane point emissions with the PRISMA spaceborne imaging spectrometer, Remote Sensing of Environment, 265, 112 671, 2021.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, https://api.semanticscholar.org/CorpusID:206594692, 2015.

Huang, G., Liu, Z., and Weinberger, K. Q.: Densely Connected Convolutional Networks, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269, https://api.semanticscholar.org/CorpusID:9433631, 2016.

Huhs, O., Borchardt, J., Krautwurst, S., Gerilowski, K., Bovensmann, H., Bösch, H., and Burrows, J. P.: Impact of stray light on greenhouse gas concentration retrievals and emission estimates as observed with the passive airborne remote sensing imager MAMAP2D-Light, Atmospheric Measurement Techniques, 19, 871–898, https://doi.org/10.5194/amt-19-871-2026, 2026.

IPCC: Technical Summary, p. 35–144, Cambridge University Press, 2023.

Jacob, D. J., Turner, A. J., Maasakkers, J. D., Sheng, J., Sun, K., Liu, X., Chance, K., Aben, I., McKeever, J., and Frankenberg, C.: Satellite observations of atmospheric methane and their value for quantifying methane emissions, Atmospheric Chemistry and Physics, 16, 14 371–14 396, https://doi.org/10.5194/acp-16-14371-2016, 2016.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E.: Adaptive mixtures of local experts, Neural computation, 3, 79–87, 1991.

Jervis, D., McKeever, J., Durak, B. O. A., Sloan, J. J., Gains, D., Varon, D. J., Ramier, A., Strupler, M., and Tarrant, E.: The GHGSat-D imaging spectrometer, Atmospheric Measurement Techniques, 14, 2127–2140, https://doi.org/10.5194/amt-14-2127-2021, 2021.

Jongaramrungruang, S., Frankenberg, C., Matheou, G., Thorpe, A. K., Thompson, D. R., Kuai, L., and Duren, R. M.: Towards accurate methane point-source quantification from high-resolution 2-D plume imagery, Atmospheric Measurement Techniques, 12, 6667–6681, https://doi.org/10.5194/amt-12-6667-2019, 2019.

Jongaramrungruang, S., Thorpe, A., Matheou, G., and Frankenberg, C.: MethaNet – An AI-driven approach to quantifying methane point-source emission from high-resolution 2-D plume imagery, Remote Sensing of Environment, 269, 112 809, https://doi.org/10.1016/j.rse.2021.112809, 2022.

Joyce, P., Ruiz Villena, C., Huang, Y., Webb, A., Gloor, M., Wagner, F. H., Chipperfield, M. P., Barrio Guilló, R., Wilson, C., and Boesch, H.: Using a deep neural network to detect methane point sources and quantify emissions from PRISMA hyperspectral satellite images, Atmospheric Measurement Techniques, 16, 2627–2640, https://doi.org/10.5194/amt-16-2627-2023, 2023.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, https://arxiv.org/abs/1412.6980, 2017.

Krings, T., Gerilowski, K., Buchwitz, M., Reuter, M., Tretner, A., Erzinger, J., Heinze, D., Pflüger, U., Burrows, J. P., and Bovensmann, H.: MAMAP – a new spectrometer system for column-averaged methane and carbon dioxide observations from aircraft: retrieval algorithm and first inversions for point source emission rates, Atmospheric Measurement Techniques, 4, 1735–1758, https://doi.org/10.5194/amt-4-1735-2011, 2011.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet classification with deep convolutional neural networks, Communications of the ACM, 60, 84 – 90, https://api.semanticscholar.org/CorpusID:195908774, 2012.

Kuhlmann, G., Henne, S., Meijer, Y., and Brunner, D.: Quantifying CO_2 emissions of power plants with CO_2 and NO_2 imaging satellites, Frontiers in Remote Sensing, 2, 689 838, 2021.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B.: Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9992–10 002, https://api.semanticscholar.org/CorpusID:232352874, 2021.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S.: A ConvNet for the 2020s, https://arxiv.org/abs/2201.03545, 2022.

Lu, X., Jacob, D. J., Zhang, Y., Shen, L., Sulprizio, M. P., Maasakkers, J. D., Varon, D. J., Qu, Z., Chen, Z., Hmiel, B., et al.: Observation-derived 2010-2019 trends in methane emissions and intensities from US oil and gas fields tied to activity metrics, Proceedings of the National Academy of Sciences, 120, e2217900 120, 2023.

Matheou, G. and Bowman, K. W.: A recycling method for the large-eddy simulation of plumes in the atmospheric boundary layer, Environmental Fluid Mechanics, 16, 69–85, 2016.

Ocko, I. B., Sun, T., Shindell, D., Oppenheimer, M., Hristov, A. N., Pacala, S. W., Mauzerall, D. L., Xu, Y., and Hamburg, S. P.: Acting rapidly to deploy readily available methane mitigation measures by sector can immediately slow global warming, Environmental Research Letters, 16, 054 042, 2021.

Ouerghi, E., Ehret, T., Facciolo, G., Meinhardt, E., Marion, R., and Morel, J.-M.: Tightening up methane plume source rate estimation in EnMAP and PRISMA images, Atmospheric Measurement Techniques, 18, 4611–4629, https://doi.org/10.5194/amt-18-4611-2025, 2025.

Plewa, T., Butz, A., Frankenberg, C., Thorpe, A. K., and Marshall, J.: Improvements of AI-driven emission estimation for point sources applied to high resolution 2-D methane-plume imagery, Remote Sensing of Environment, 331, 115 002, https://doi.org/10.1016/j.rse.2025.115002, 2025.

Radman, A., Mahdianpari, M., Varon, D. J., and Mohammadimanesh, F.: S2MetNet: A novel dataset and deep learning benchmark for methane point source quantification using Sentinel-2 satellite imagery, Remote Sensing of Environment, 295, 113 708, 2023.

Roger, J., Irakulis-Loitxate, I., Valverde, A., Gorroño, J., Chabrillat, S., Brell, M., and Guanter, L.: High-Resolution Methane Mapping With the EnMAP Satellite Imaging Spectroscopy Mission, IEEE Transactions on Geoscience and Remote Sensing, 62, 1–12, https://doi.org/10.1109/TGRS.2024.3352403, 2024.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. S., Berg, A. C., and Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision, 115, 211 – 252, https://api.semanticscholar.org/CorpusID:2930547, 2014.

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., Raymond, P. A., Dlugokencky, E. J., Houweling, S., Patra, P. K., Ciais, P., Arora, V. K., Bastviken, D., Bergamaschi, P., Blake, D. R., Brailsford, G., Bruhwiler, L., Carlson, K. M., Carrol, M., Castaldi, S., Chandra, N., Crevoisier, C., Crill, P. M., Covey, K., Curry, C. L., Etiope, G., Frankenberg, C., Gedney, N., Hegglin, M. I., Höglund-Isaksson, L., Hugelius, G., Ishizawa, M., Ito, A., Janssens-Maenhout, G., Jensen, K. M., Joos, F., Kleinen, T., Krummel, P. B., Langenfelds, R. L., Laruelle, G. G., Liu, L., Machida, T., Maksyutov, S., McDonald, K. C., McNorton, J., Miller, P. A., Melton, J. R., Morino, I., Müller, J., Murguia-Flores, F., Naik, V., Niwa, Y., Noce, S., O'Doherty, S., Parker, R. J., Peng, C., Peng, S., Peters, G. P., Prigent, C., Prinn, R., Ramonet, M., Regnier, P., Riley, W. J., Rosentreter, J. A., Segers, A., Simpson, I. J., Shi, H., Smith, S. J., Steele, L. P., Thornton, B. F., Tian, H., Tohjima, Y., Tubiello, F. N., Tsuruta, A., Viovy, N., Voulgarakis, A., Weber, T. S., van Weele, M., van der Werf, G. R., Weiss, R. F., Worthy, D., Wunch, D., Yin, Y., Yoshida, Y., Zhang, W., Zhang, Z., Zhao, Y., Zheng, B., Zhu, Q., Zhu, Q., and Zhuang, Q.: The Global Methane Budget 2000–2017, Earth System Science Data, 12, 1561–1623, https://doi.org/10.5194/essd-12-1561-2020, 2020.

Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., and Dean, J.: Outrageously large neural networks: The sparsely-gated mixture-of-experts layer, arXiv preprint arXiv:1701.06538, 2017.

Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, CoRR, abs/1409.1556, https://api.semanticscholar.org/CorpusID:14124313, 2014.

Strandgren, J., Krutz, D., Wilzewski, J., Paproth, C., Sebastian, I., Gurney, K. R., Liang, J., Roiger, A., and Butz, A.: Towards spaceborne monitoring of localized CO_2 emissions: an instrument concept and first performance assessment, Atmospheric Measurement Techniques, 13, 2887–2904, https://doi.org/10.5194/amt-13-2887-2020, 2020.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z.: Rethinking the Inception Architecture for Computer Vision, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818–2826, https://api.semanticscholar.org/CorpusID:206593880, 2015.

Tan, M. and Le, Q. V.: EfficientNetV2: Smaller Models and Faster Training, in: International Conference on Machine Learning, https://api.semanticscholar.org/CorpusID:232478903, 2021.

Thorpe, A. K., Frankenberg, C., and Roberts, D. A.: Retrieval techniques for airborne imaging of methane concentrations using high spatial and moderate spectral resolution: application to AVIRIS, Atmospheric Measurement Techniques, 7, 491–506, https://doi.org/10.5194/amt-7-491-2014, 2014.

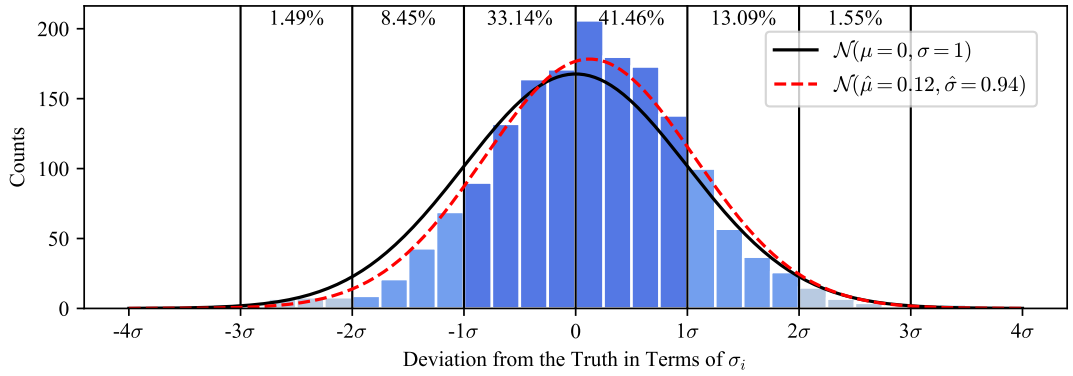
Thorpe, A. K., Frankenberg, C., Thompson, D. R., Duren, R. M., Aubrey, A. D., Bue, B. D., Green, R. O., Gerilowski, K., Krings, T., Borchardt, J., Kort, E. A., Sweeney, C., Conley, S., Roberts, D. A., and Dennison, P. E.: Airborne DOAS retrievals of methane, carbon dioxide, and water vapor concentrations at high spatial resolution: application to AVIRIS-NG, Atmospheric Measurement Techniques, 10, 3833–3850, https://doi.org/10.5194/amt-10-3833-2017, 2017.

Thorpe, A. K., Green, R. O., Thompson, D. R., Brodrick, P. G., Chapman, J. W., Elder, C. D., Irakulis-Loitxate, I., Cusworth, D. H., Ayasse, A. K., Duren, R. M., Frankenberg, C., Guanter, L., Worden, J. R., Dennison, P. E., Roberts, D. A., Chadwick, K. D., Eastwood, M. L., Fahlen, J. E., and Miller, C. E.: Attribution of individual methane and carbon dioxide emission sources using EMIT observations from space, Science Advances, 9, eadh2391, https://doi.org/10.1126/sciadv.adh2391, 2023.

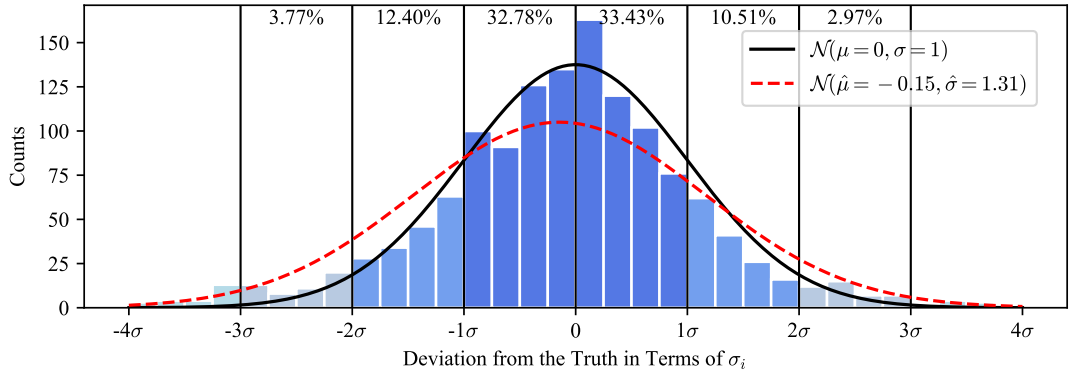
Varon, D. J., Jacob, D. J., McKeever, J., Jervis, D., Durak, B. O. A., Xia, Y., and Huang, Y.: Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes, Atmospheric Measurement Techniques, 11, 5673–5686, https://doi.org/10.5194/amt-11-5673-2018, 2018.

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., De Vries, J., Otter, G., Claas, J., Eskes, H., De Haan, J., Kleipool, Q., et al.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, Remote Sensing of Environment, 120, 70–83, 2012.

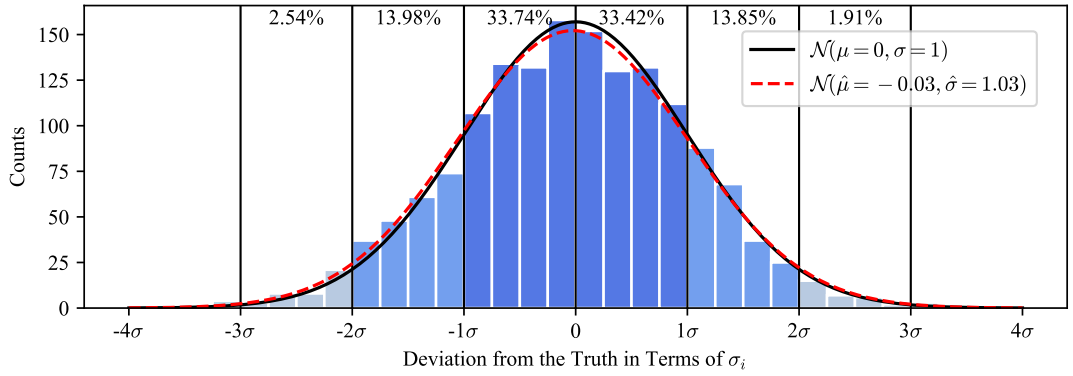
Zavala-Araiza, D., Alvarez, R. A., Lyon, D. R., Allen, D. T., Marchese, A. J., Zimmerle, D. J., and Hamburg, S. P.: Super-emitters in natural gas infrastructure are caused by abnormal process conditions, Nature communications, 8, 14 012, 2017.



(a) Filtered wind speed ensemble for 1 m s^{-1}

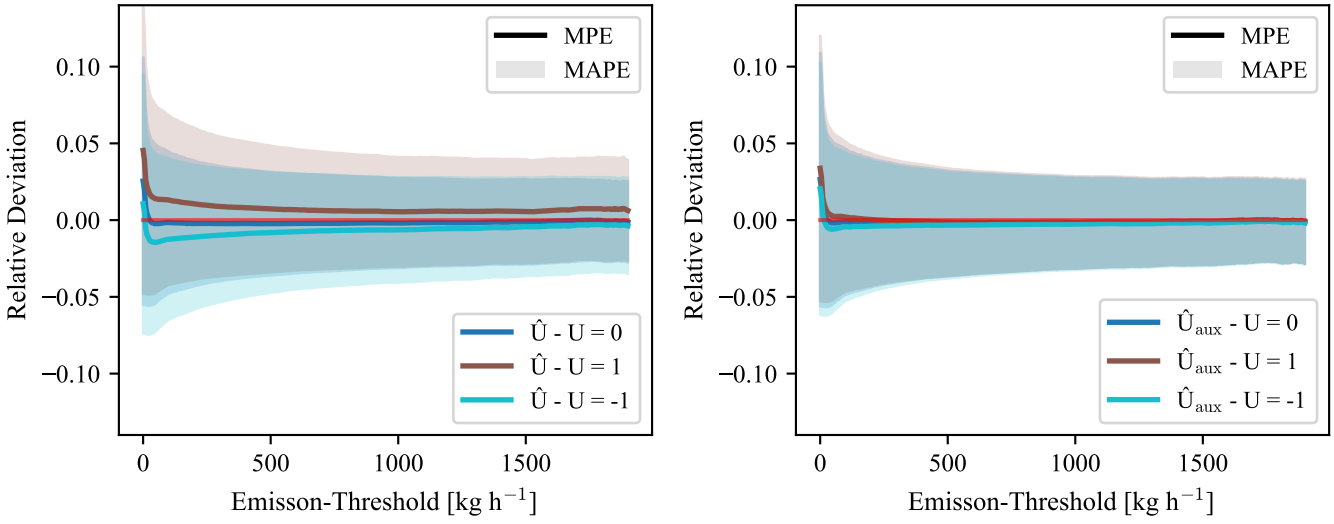


(b) Filtered wind speed ensemble for 8 m s^{-1}



(c) Filtered wind speed ensemble for 9 m s^{-1}

Figure 9. Deviations of the predicted from the true emission rates with respect to the corresponding estimated uncertainty for each scene for different wind speed ensembles, filtered for wind speed conditions matching within $\pm$ 1 m s^{-1}. The ideal distributions are depicted in black and the best fits to the data in red.



(a) Only external data

(b) Auxiliary data only for strong deviations

Figure 10. Mean percentage error and mean absolute percentage error for all scenes above a certain emission threshold for different scenarios. One the left, all scenes are evaluated either at the correct wind speed assignment or at an assignment that is $\pm 1\,\mathrm{m\,s^{-1}}$, on the right side all the scenes that show a deviation larger than $\pm 1\,\mathrm{m\,s^{-1}}$ get replaced either by the correct wind speed scenario or a scenario that is $\pm 1\,\mathrm{m\,s^{-1}}$ correct.

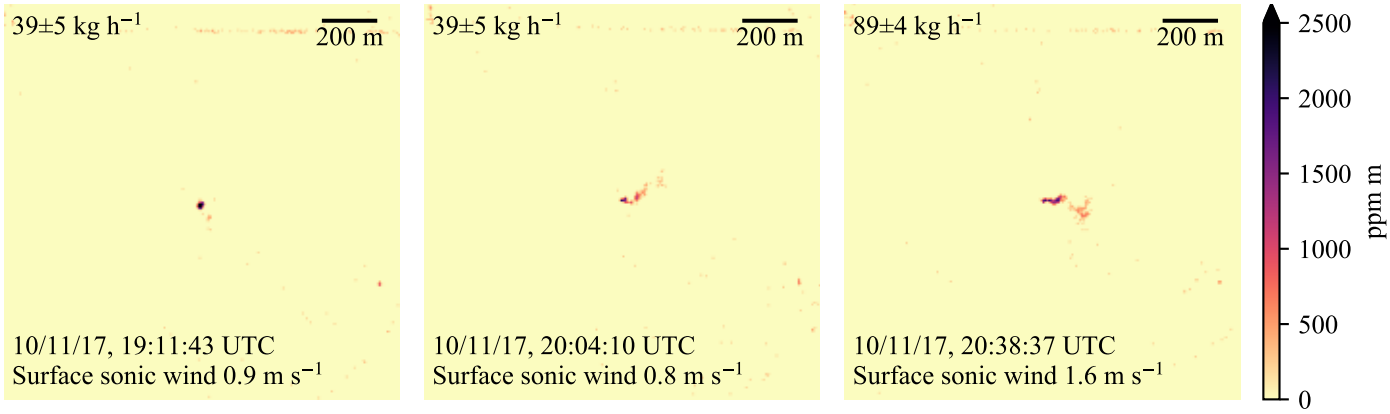


Figure 11. Controlled release experiments observed with AVIRIS-NG from a natural gas pipeline on 11 October 2017 located at $34.76°, -117.28°$. The expected emission rate is shown in the top left corner of the scene and the measured wind speed during the overpass (Jongaramrungruang et al., 2019, 2022) is depicted at the bottom of the scene alongside the acquisition time and date.