Enabling Fast Greenhouse Gas Emissions Inference from Satellites with GATES: a Graph-Neural-Network Atmospheric Transport Emulation System

Elena Fillola[1,2*], Raul Santos-Rodriguez[2], Rachel Tunnicliffe[1], Jeffrey N. Clark[2], Nawid Keshtmand[1], Anita Ganesan[3], Matthew Rigby[1]

[1]School of Chemistry, University of Bristol, Bristol, UK
[2]School of Engineering Mathematics and Technology, University of Bristol, Bristol, UK
[3]School of Geographical Sciences, University of Bristol, Bristol, UK

Correspondence to: Elena Fillola (elena.fillolamayoral@bristol.ac.uk)

Abstract. Atmospheric observation-based "inverse" greenhouse gas flux estimates are increasingly important to evaluate national inventories. A dramatic improvement in "top-down" flux inference is expected in the coming years due to the rapidly growing number of measurements from space. However, many well-established inverse modelling techniques face significant computational challenges scaling to modern satellite datasets, particularly those that rely on Lagrangian Particle Dispersion Models (LPDM) to simulate atmospheric transport. Here, we introduce GATES (Graph-Neural-Network Atmospheric Transport Emulation System), a data-driven LPDM emulator which outputs source-receptor relationships ("footprints") using only meteorology and surface data as inputs, approximately 1000x times faster than an LPDM. We demonstrate GATES's skill in estimating footprints over South America and integrate it into an emissions estimation pipeline, evaluating Brazil's methane emissions using GOSAT (Greenhouse Gases Observing Satellite) observations for 2016 and 2018 and finding emissions that are consistent in space and time with the physics-driven estimate. This work highlights the potential of machine learning-based emulators like GATES to overcome a key bottleneck in large-scale, satellite-based inverse modeling, accelerating greenhouse gas emissions estimation and enabling timely, improved evaluations of national GHG inventories.

1 Introduction

Reducing greenhouse gas (GHG) emissions is essential for mitigating climate change, with carbon dioxide (CO_2) and methane (CH_4) being two of the most significant contributors. International agreements such as the Global Methane Pledge, with over 150 participating countries, reflect the commitment to address this challenge, by aiming for a 30% reduction in emissions by 2030, compared to 2020 levels (European Commission and United States of America, 2021). However, significant uncertainties persist in the inventory-based ("bottom-up") reports of national methane emissions, which will be used to evaluate progress (Saunois et al., 2020, 2025). "Top-down" or "inverse" estimates of GHG emissions can be used to evaluate and improve these self-reported national emissions inventories and are therefore seen as a valuable tool for supporting international climate agreements (e.g., (Leip et al., 2018)). These methods use atmospheric concentration observations to quantify surface fluxes, with atmospheric chemical transport models providing the link between these two quantities.

Until recently, top-down studies primarily relied on in situ atmospheric observations from individual stations or networks (e.g., (Bergamaschi et al., 2018). Increasingly, CH_4 and CO_2 flux estimates are being derived using satellite data, reflecting the expansion of space-based instruments (e.g., (Ganesan et al., 2017; Scarpelli et al., 2022; Tunnicliffe et al., 2020; Western et al., 2021; Worden et al., 2023) , which can provide valuable insights into regions previously under-sampled by in situ data. National emissions estimates have been derived using data from instruments such as GOSAT (Greenhouse gases Observing SATellite , launched 2009 (Parker et al., 2020)) using regional (Ganesan et al., 2017; Tunnicliffe et al., 2020; Western et al., 2021) or global (Alexe et al., 2015; Feng et al., 2023; Maasakkers et al., 2019) chemical transport models. Current global instruments, like TROPOMI (Tropospheric Monitoring Instrument, launched 2017 (Veefkind et al., 2012)) have over 100 times the observation density of GOSAT, and recently launched and upcoming satellites (e.g. CO2M, MethaneSAT) will continue to grow the volume of available observations (Jacob et al., 2022). Higher data density increases the spatial and

temporal resolution at which fluxes can be inferred providing the opportunity to better understand the processes driving observed changes in global atmospheric CO_2 and CH_4 abundance.

This rapid growth in data volume, from in situ-based datasets (~1000s of observations per month in a continental region) to satellite-based (~10,000s to 100,000s of observations per month in a continental region), is leading to scaling issues for many
45   traditional approaches to flux inference. One of the most common families of top-down methods for national scale emissions estimation employs Lagrangian Particle Dispersion Models (LPDMs), which use archived meteorological data to simulate the movement of hypothetical gas particles backwards in time from a location to the surface in the surrounding region. The advantage of using LPDMs for GHG flux inference is that LPDMs directly calculate the sensitivity, or "footprint" (Fig. 1a), of a mole fraction measurement to upwind fluxes. This contrasts with Eulerian models, which simulate concentrations in a 3D
50   atmospheric grid rather than using individual particles, and do not directly calculate source–receptor relationships. To be useful in top-down inference, Eulerian model outputs require additional processing to derive sensitivities of observations to fluxes, such as performing ensembles of perturbed flux runs (e.g., (Baker, 2006a; Bousquet, 2000; Peters et al., 2005)) or deriving adjoint model code (e.g., (Baker, 2006b; Kaminski et al., 1999). However, a major disadvantage of LPDMs is that they require one run for each observation (several core-minutes, depending on set-up), leading to high computational burdens for large
55   datasets. Studies show that increasing the number of simulated particles and extending their simulation time reduces statistical errors in the transport model, but these improvements also come at the expense of even greater resource demands (Pisso et al., 2019; Vojta et al., 2022). Although LPDMs are easily parallelizable with high performance clusters allowing for multiple simulations to run simultaneously, the overall computational cost remains substantial and will likely be a barrier in the use of LPDM-based methods for applications involving planned space-based GHG measurements.

60   The satellite community has recognized computational capacity as a key limitation in applying greenhouse gas measurements and inverse modelling to support policy (Joint CEOS-CGMS Working Group on Climate - Greenhouse Gas Task Team., 2024).. Two recent studies that use Lagrangian methods for TROPOMI observations over Alaska and Siberia propose methods to coarsen (Thompson et al., 2025) or subsample (Ward et al., 2025) the dense observational dataset, so that it is feasible to run LPDMs. Several recent studies have proposed methods for interpolating or emulating footprints (see below) but none have
65   addressed the challenging problem of emulating satellite footprints at continental scales. In this study, we present GATES (Graph-Neural-Network Atmospheric Transport Emulation System), a new method for footprint emulation that addresses previous limitations, offering high computational efficiency for satellite data whilst retaining sufficient accuracy over large scales for regional flux inference. Section 2 provides background on LPDMs and previous emulation approaches, and Section 3 presents the architecture of the machine learning model. In Section 4 we describe the generation of LPDM footprints in our
70   case study region, South America, and detail the inputs, dataset and training approach. In Section 5 we show and evaluate the outputs of GATES, and most importantly we demonstrate its application in a previously studied top-down inference pipeline to estimate Brazil's methane emissions.

2 Background and problem statement

2.1 Lagrangian Particle Dispersion Models

75   To calculate the sensitivity of a mole fraction measurement to surface fluxes using LPDMs, thousands of virtual particles are released from a set of 3D coordinates (latitude, longitude, height) and transported backwards in time for a number of days. The model records whenever these particles are near the surface (within 40m in the simulations used here) throughout the whole time period, creating an aggregated 2D "influence footprint" for each observation that indicates the contribution of a unit surface flux at a particular location to the observed mole fraction. When applied to in situ measurements, the LPDM is
80   initialized using the site coordinates and the inlet height. Satellites, on the other hand, take column-averaged measurements of GHG mole fractions, so instead, a number of 3D releases are run from different heights in the atmosphere, and then averaged using a kernel (e.g. Ganesan et al. (2017)). This means that LPDM runs for satellite footprints are more computationally expensive than for in-situ footprints.

85 Here, we emulate the UK Met Office LPDM NAME (Numerical Atmospheric-dispersion Modelling Environment (Jones et al., 2007)), with a similar setup to those used by Tunnicliffe et al. (2020) and Ganesan et al. (2017), who run NAME for valid GOSAT observations to infer methane emissions. They perform NAME simulations for 20 vertical levels, releasing 1000 particles per hour for levels 1–8 and at 100 per hour for levels 8–17. NAME cannot emulate dispersion for the upper three column levels (18-20) and therefore the prior column mole fractions are used. The particles are simulated travelling backwards

90 in time for 30 days from the location and time of each satellite measurement, recording interactions with the surface whenever they are below 40m above ground level. The calculated sensitivity maps for each vertical level, output at a resolution of 0.352×0.234°, are height-averaged into a single 2D footprint, weighted by the corresponding GOSAT averaging kernel and pressure weight. See Ganesan et al. (2017) for more details on the set-up. Other examples of well-known LPDMs that have been applied to similar applications are the FLEXible PARTicle Dispersion Model (FLEXPART (Pisso et al., 2019)), the

95 Hybrid Single-Particle Lagrangian Integrated Trajectory model (HYSPLIT, (Stein et al., 2015)) and the Stochastic Time-Inverted Lagrangian Transport Model (STILT, (Fasoli et al., 2018)).

2.2 Meteorology

LPDMs are usually driven by a meteorological reanalysis product. Here, our NAME simulations use the Met Office Unified

100 Model (UM) global analysis fields to simulate particle dispersion. The global model has a resolution of 3 hours and 25 km up to July 2014, 17 km from then until July 2017 and 12 km thereafter, with 59 vertical levels up to 29 km (Met Office, 2025a, b). NAME also uses static fields (i.e. not time dependent) of surface characteristics, including orography, a sea-land mask, and a land surface type map with nine categories (including different types of trees and grasslands, urban, bare soil and ice (Essery et al., 2001)), at a resolution of 0.156°×0.234°.

105

2.3 Problem statement

Summarised mathematically, an LPDM f produces a footprint y_t^φ by modelling backwards particle dispersion from a set of coordinates φ at time t. For this, it uses static information ω including topography and land-use maps, and a time series of 3D meteorological features, starting at time t (x^t) and extending backwards in time from t for a number of steps N with separation

110 Δt. The meteorological input time series can therefore be defined as $x^{t:t-N\Delta t} = [x^t, x^{t-\Delta t}, \dots, x^{t-N\Delta t}]$, and the LPDM summarized as $y_t^\varphi = f(\varphi, x^{t:t-N\Delta t}, \omega)$. The LPDM setup described here uses $\Delta t = 3h$, and $N = 240$, which equates up to tracking particles backwards in time for 30 days.

Our goal is to develop an efficient emulator Φ that can accurately recreate footprints produced by the LPDM in orders of

115 magnitude less time and potentially using only a subset of the full LPDM inputs. Thus, the emulator predicts a footprint as $\hat{y}_t^\varphi = \Phi(\varphi, \tilde{x}, \tilde{\omega})$ where $\tilde{x} \subseteq x^{t:t-N\Delta t}$ and $\tilde{\omega} \subseteq \omega$.

2.4 Previous ML approaches

The urgent need for scalable LPDM-based inverse methods was previously addressed by Roten et al. (2021), Fasoli et al.

120 (2018), Cartwright et al. (2021) and others, who derived modest gains in computational efficiency through interpolation-based approaches. More complex machine learning (ML) methods such as those developed by Fillola et al. (2023) and He et al., (2024) introduced proof-of-concept emulators of LPDM footprints for surface sites using meteorological fields as inputs. One of these studies (Fillola et al., 2023) used a set of individual regressors to emulate the footprint value at each grid-cell for in-situ observation towers, but only estimated these values for part of the footprint, within ~100 km of the measurement site.

125 Therefore, to perform flux inversions, the emulated footprints had to be nested within a low-resolution LPDM simulation. The emulator developed in ((He et al., 2024) focuses on emulating high-resolution in-situ footprints (400×400 km^2 domain at 1 km resolution) for an urban sensor network (Dadheech et al., 2024). To constrain their convolutional neural network architecture, one of the key inputs was a Gaussian plume, a simplified footprint approximation calculated using the meteorology at the measurement site. While this approach is effective for local-scale dispersion, it is unlikely to be suitable for national or

130 continental transport (~1000–5000 km) because the assumptions underpinning the Gaussian plume model (such as constant wind speed and homogeneous atmospheric conditions) only hold over ~10km spatial scales.
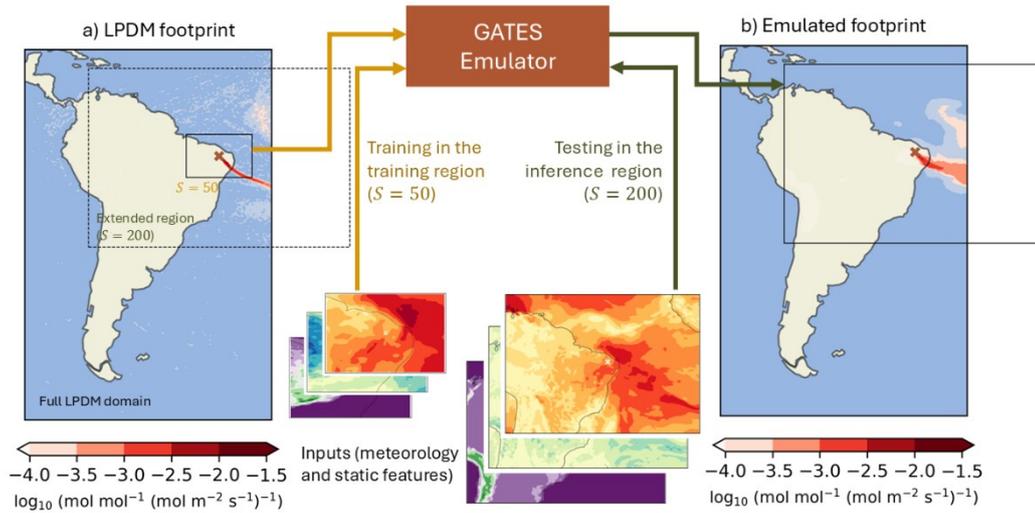
3 Model Architecture

The Graph-Neural-Network Atmospheric Transport Emulation System (GATES) is a ML-driven emulator of regional LPDM-derived satellite observation footprints using meteorological and topographical inputs. The architecture of GATES builds on that of Keisler, (2022) and Deepmind's Graphcast (Lam et al., 2023), who developed weather forecasting models using Graph Neural Networks (GNNs) with an encoder-processor-decoder structure. This architecture achieved breakthroughs in deterministic weather forecasting, sometimes outperforming the reference physical models (Lam et al., 2023), as well as effectively learning complex systems based on Partial Differential Equations (PDEs), like fluid dynamics and diffusion (Pfaff et al., 2020; Sanchez-Gonzalez et al., 2020). Many of these systems are designed to learn a one-step forward model, predicting system quantities at time $t+1$ given the current state (and often previous states too). Unlike these previous studies, GATES is trained to return time-integrated 2D GHG measurement footprints without requiring information about the intermediate model timesteps.

As inputs, GATES takes a subset of the meteorology that drives the LPDM, reducing input/output overhead and memory requirements compared to running the LPDM: we provide a number of snapshots of the state of the atmosphere at the time of the satellite measurement and before, including variables like wind direction, wind speed and planetary boundary layer height, as a latitude-longitude grid and at selected vertical levels. These inputs are complemented with time-independent variables, including the latitude-longitude coordinates of each grid-cell , its distance from the measurement, the surface height above sea level, and the main landcover type.

Graph Neural Networks represent data as a graph, composed of nodes and connecting edges. In GATES, each coordinate in the latitude-longitude grid is treated as a node, with the inputs at that location considered its attributes. The encoder-processor-decoder architecture means that the inputs on a latitude-longitude grid are first translated ("encoded") into an abstract space, "processed" iteratively to spread the information, and then "decoded" back into the original latitude-longitude grid, making predictions for the footprint value at each grid-cell. Figure 1 shows a schematic of the architecture of GATES. In the encoder (Fig. 1c), the input grid is encoded in a lower-resolution regular triangular mesh, which acts as an internal abstract feature space. This mesh, built with the h3 library (Uber Technologies Inc, 2024), divides the domain into hexagons with an average area of 1770km^2, and with mesh nodes placed ~40km from each other. We place edges between each node in the lat-lon grid and its closest node in the mesh, so that each mesh node receives a Multi-Layer Perceptron (MLP) encoding of the distance-weighted mean of the local features. MLPs are simple feedforward neural networks composed of multiple layers of linear transformations and non-linear transformation functions, which here are used to update node and edge features based on local information. In the processor (Fig. 1d), each node in the mesh is connected to its six adjacent nodes, and the whole mesh is updated in multiple message-passing rounds. Each message-passing GNN block (Battaglia et al., 2018) first uses an MLP to update the features of the edges in the mesh using information from the adjacent nodes, and then another MLP to update each node based on its adjacent features. In the decoder (Fig. 1g), a final MLP maps the mesh-node features back to the original latitude-longitude grid, outputting the footprint (Fig. 1e).

GATES is inherently flexible because it operates through local, node-wise computations. At every stage (encoding, processing and decoding) the model learns to update each node based only on its neighbours' features, rather than relying on a regular grid or fixed input dimensions. Traditional architectures such as Convolutional Neural Networks (CNNs), on the other hand, require inputs to be defined on uniform, fixed-sized grids, where emulation depends on the resolution and dimensions of the training domain. With GATES, training and inference can be done at different resolutions, geometries or domain sizes, without needing to retrain or modify the architecture.

Figure 1: GATES model schematic and architecture.
Top: (a) Example of an LPDM-generated footprint (~20 mins to generate on a single CPU) showing the sensitivity of a satellite measurement to fluxes from the surrounding domain. (b) Corresponding GATES-emulated footprint (~1.5s to generate). Both (a) and (b) are shown over the full LPDM domain and resolution (0.352°×0.234°, ≈ 33×25km in mid-latitudes) used in (Tunnicliffe et al., 2020). In (a), the solid-line box corresponds to the training region, a square of size 50×50 grid-cells centered around the coordinates of the satellite measurement. The solid-line box in (b) corresponds to the inference region, a 200x200 grid-cell square over which footprints were emulated. When this area escapes the original footprint domain, as above, the added space is filled with zeros and ignored in evaluation (here shown as whitespace inside the boxes). Bottom: Architecture of the graph and emulator, demonstrated for a square domain of size 5×5. GATES is composed of (c) an encoder, (d) a processor, and (e) a decoder. The black solid arrows indicate the flow of information. The inputs are represented as node attributes on a latitude-longitude grid (crosses) centered around the measurement point (red cross). (c) In the encoder, local regions of the inputs are mapped into nodes in the abstract feature space (yellow hexagons), arranged as an triangular mesh. Each mesh-node is connected with edges to its six neighbors (yellow lines). (d) In the processor, each mesh-node is updated using message-passing from the neighboring nodes. Four independent message-passing blocks are learned. (e) The decoder maps the latent space back to the original latitude-longitude grid of the same resolution as the inputs, outputting the footprint value at each node.

5

4 GATES in practice: Emulating GOSAT footprints over South America

190

To demonstrate the model, we train GATES to emulate NAME footprints for GOSAT observations over South America, as presented in (Tunnicliffe et al., 2020). Tunnicliffe et al. produce NAME footprints for GOSAT observations that pass the quality threshold over an area defined by 35.8° S to 7.3° N and from 76.0 to 32.8° W. They produced observation for both nadir measurements (over the land) and glint measurements (over the ocean), but, in Tunnicliffe et al, they use only nadir measurements within the inversions to avoid any biases between the two. GATES is trained and validated using data from 2014 to 2016, both over the land and the ocean, and emissions are derived for 2016 and 2018, using only land-based measurements following the reference study. The year 2018 was chosen to demonstrate the performance of the GATES-based inversion as it is separated in time from the data used in the training and validation set. The footprints aggregate the surface interactions of particles transported for the 30 days leading up to the satellite measurement at a resolution of 0.352° longitude × 0.234° latitude (≈ 33×25km in mid latitudes), constrained to a domain that covers the whole South American continent (60.98° S to 22.32° N and 91.33° W to 24.8° W with a grid of size 190 × 357). Each footprint takes about ~20 core-minutes to be generated on a high-performance cluster.

4.1 Inputs: meteorology and topography

This version of the emulator takes as main inputs nine physical variables at seven vertical levels in the atmosphere (see Table 1 for all the variables and levels used). The meteorology is interpolated linearly from its 3-hourly resolution t, as well as any required previous timesteps up to $t - \widetilde{N}\widetilde{\Delta}t$. In the application described here, $\widetilde{N} = 2$ and $\widetilde{\Delta}t = 6h$, so that the meteorological input to each footprint is composed of the weather state at t, six hours and twelve hours before. This is a small subset of the full NAME input meteorology, which usually requires ~15 meteorological variables at 50 atmospheric levels at over 200 timesteps. We also provide the model with information from the static fields NAME uses to describe topography (orography and landcover type), as well as two variables with the x- and y- coordinates of each grid-cell with respect to the measurement coordinates (so that the grid-cell with the satellite observation is [0,0]) and a variable with the Euclidean distance from each grid-cell to the measurement coordinates. Each input is a 2D field of the same resolution and domain as the corresponding footprint. The vertical levels and the number of timesteps were decided through tuning, where a small range of hand-crafted configurations were tested.

215

Variable name	Description	Vertical levels
x-wind		
y-wind		
Upward air velocity		100m, 660m, 1.7km, 3.2km, 6.4km, 12km, 18km
Air temperature	At t=t$_{obs}$, t=t$_{obs-6h}$ and t=t$_{obs-12h}$	
Air pressure		
Wind speed		100m, 6.4km, 18km
Wind direction		
Boundary layer thickness		Surface variables (0m)
Surface air pressure		
Latitude	Not time-dependent	
Longitude	Not time-dependent	
Distance from measurement point	Not time-dependent, Euclidean distance in x-y grid (not on earth surface)	
x-coordinate	Not time-dependent	
y- coordinate	Not time-dependent	
Height above sea level	Not time-dependent	
Main landcover type	Not time-dependent	

Table 1 – Input variables provided to the emulator. In total, there are (2 atmospheric surface variables + 7 atmospheric variables × 7 levels) × 3 timesteps + 2 topography variables + 5 constant variables = 160 input features.

4.2 Preparing the dataset

220   GATES is designed to be LPDM domain agnostic: instead of emulating footprints over the whole LPDM domain, which tends to be sparse, GATES emulates footprints on a square domain of $S \times S$ grid cells centred around the measurement coordinates φ. The meteorology and static fields are cropped to the same area for each corresponding footprint.

For this South America application, GATES is first trained on a square (in latitude/longitude space) of size S=50 grid cells
225   centered around the satellite measurement. Subsequently, footprints are emulated over a square of S=200 grid-cells (~6000×5000 km; see the training and inference regions in Fig 1, panels a and b). We find that training in the smaller domain is not only more computationally efficient, but that the model performs better than training on the S=200 domain, likely because the wider domain is mostly sparse. We choose to define our inference region as S=200 as our preliminary tests suggest this size of domain doesn't introduce substantial error in the simulated mole fractions, compared to simulations using a larger
230   region. When using the emulated footprints in the inference emissions pipeline, we fill the rest of the full LPDM domain with zeros.

We split our dataset into training, validation, and test sets using distinct time periods, to imitate deployment conditions and to avoid the high autocorrelations in meteorological fields from polluting the evaluation results. We use the first two years to
235   train the model (2014-2015, 11165 samples), validate on January, February and March of 2016 (4314 footprints), and evaluate the outputs for the remaining months of 2016, as well as the whole of 2018 (36671 footprints, 16945 from 2016 and 15412 from 2018). The validation set was used to tune model hyperparameters, and to calibrate the bias correction. Figure A1 shows the spatial and seasonal distribution of the datasets used.

240
4.3 Training details

4.3.1 Data normalization

During training, the non-zero values in the footprints are log-transformed and shifted by the minimum value in the training dataset, so that all non-zero values in the original footprints y remain strictly positive in the transformed footprints y'. The
245   zero values are maintained as such. This helps tackle the footprints' strong exponential distributions, with values decaying quickly with distance from the measurement point. The emulated footprints on the test set are returned to the original space for evaluation. All inputs are standardized (zero mean and unit variance). Each atmospheric variable is standardized per atmospheric level (where relevant) and across all different input timesteps.

4.3.2 Model parameters

250   All neural networks in the GNN are MLPs with two hidden layers of size 16 and an output layer size of 64, except the decoder MLP, which has only one hidden layer and an output size of 1 (as a single value is predicted at each node). Here, size refers to the number of neurons in each layer. The size of the layers and the number of the hidden layers was decided through tuning, testing a number of configurations. Deeper networks or with higher number of nodes led to the model over-fitting to the training data, whereas shallower networks achieved low skill as they were unable to learn complex dispersion patterns. All
255   MLPs use Relu (Nair et al., 2010) as the activation function and are followed by a LayerNorm (Ba et al., 2016) layer (except the decoder's MLP, which outputs the footprint value).

4.3.3 Loss function

The loss function L weights mean squared error (MSE) by the footprint value, which applies stronger penalties for mispredicting areas with higher relevance. It first calculates the pixel-by-pixel squared error between the predictions and the true footprints (both in the transformed space, i.e. $\hat{y}'$ and y' respectively), and then weights it by a linear transformation $g()$ of the original footprint y, before taking the mean across the footprint. For each batch, we average the loss from all the footprints uniformly.

$$L = \ell(\hat{y}', y') = \frac{1}{S^2}\sum_{i=1}^{S}\sum_{j=1}^{S} g(y(i,j)) \times (\hat{y}'(i,j) - y'(i,j))^2$$

Where

$$g(y) = wy + a$$

Here, we use $w = 1000$ and $a = 0.5$. This training objective was minimized with the ADAMW (Lam et al., 2023; Loshchilov and Hutter, 2019) optimizer and a learning rate of 5×10^{-5}, which controls the size of the parameter updates during optimization. We use a batch size of 5, which determines how many samples get processed together in each update during training, regularising the learning.

4.4 Thresholding and bias correcting the footprints

ML regression tasks, which predict continuous values, often struggle to predict exactly zero, predicting instead very small values. The footprints, however, are sparse, with large regions of areas with no sensitivity (i.e. zeros). Therefore, following inference, the near-zero background values below a calculated threshold are set to zero, imitating approaches from precipitation modelling (Schmidli et al., 2006). The footprints are also bias-corrected by applying quantile mapping (Met Office, 2018). This bias correction is necessary to avoid the model's tendency to underpredict, likely caused by the sparsity of the outputs. The calibration of the threshold and the quantile mapping are both done using the validation set, transformed back to the original data space. GATES is not retrained after thresholding and bias correction: these two statistical steps are applied as post-processing of the already emulated footprints. Figure B1 in Appendix B illustrates the steps for a particular footprint.

4.4.1 Output thresholding

A "footprint threshold" is defined on the validation set such that the threshold exceedance frequency across all the grid-cells in the emulated footprints matches the above-zero frequency in the LPDM-generated footprints (Schmidli et al., 2006), so that there is the same number of non-zero values in both validation datasets. The threshold is calculated to be 1.76×10^{-5}, which is consistent with just removing small background values. In the test dataset, all values under this threshold are set to zero.

4.4.2 Bias correction

After thresholding, the distribution is corrected using quantile mapping. This technique maps the cumulative distribution function (CDF) of the emulated data to that of the true data, applying a transformation function $T(x)$. We define $T(x)$ by dividing the by splitting the CDF of the true data in the validation set in 5000 quantiles, and the CDF of the emulated data into the same number of quantiles, and mapping the two. After $T(x)$ has been defined on the validation set, it is applied to the test set. Figure B2 in Appendix B shows the CDF of the footprint values in the test set before and after the transformation, with an improved fit and reduced bias.

## 4.5 Hardware and time requirements

We train and run the model on a 32GB NVIDIA V100 GPU. It takes 10h to fully train, including loading all the data and model into memory. For emulation, each footprint takes around 1 second to produce on the same V100 GPU (after an initial overhead to load the model and the input data), whereas a single footprint LPDM footprint takes ~20 minutes. GATES can therefore generate footprints 1000x faster than the physics-based simulation.

Integrating the emulator in the emissions inference pipeline presents a significant speed-up. Generating footprints for a regional monthly inversion as shown here (~1,000 footprints) would take ~20 minutes with GATES, but 14 core-days with the LPDM.The inversion algorithm itself, described in Section 5.3, combines the observations, footprints and prior emissions to output updated emission fluxes takes 6–10 hours on a CPU, depending on the number of data points assimilated We use a computationally expensive hierarchical Bayesian Markov Chain Monte Carlo inverse method, as outlined in Ganesan et al. (2014).
 GATES therefore shifts the primary computational bottleneck to the inversion algorithm, which now dominates the runtime. Substantially higher overall inversion speed-ups would be obtained if a less expensive inverse method was used. As noted in (Dadheech et al., 2024), the use of an efficient emulator has the added potential advantage that footprints do not need to be archived and can instead be produced on demand, substantially reducing storage requirements. Further speed-up to GATES is expected with software optimization in the future.

5 Evaluating GATES performance: footprints, predicted concentrations and inferred emissions

In this section we use a range of tests to evaluate the performance of the emulated footprints at three stages of the inversion pipeline: 1) metrics to compare the emulated footprints to the LPDM, 2) comparison of simulated methane column mole fractions over South America, and 3) comparison of Brazil's inferred methane emissions at monthly resolution for April-December 2016 and all of 2018. Figure 2 shows four examples of LPDM footprints and their emulated counterparts, as well as the emulated methane mole fractions for those days. We show that the monthly flux estimates using GATES are comparable to those obtained using LPDM footprints at national and sub-national scales.
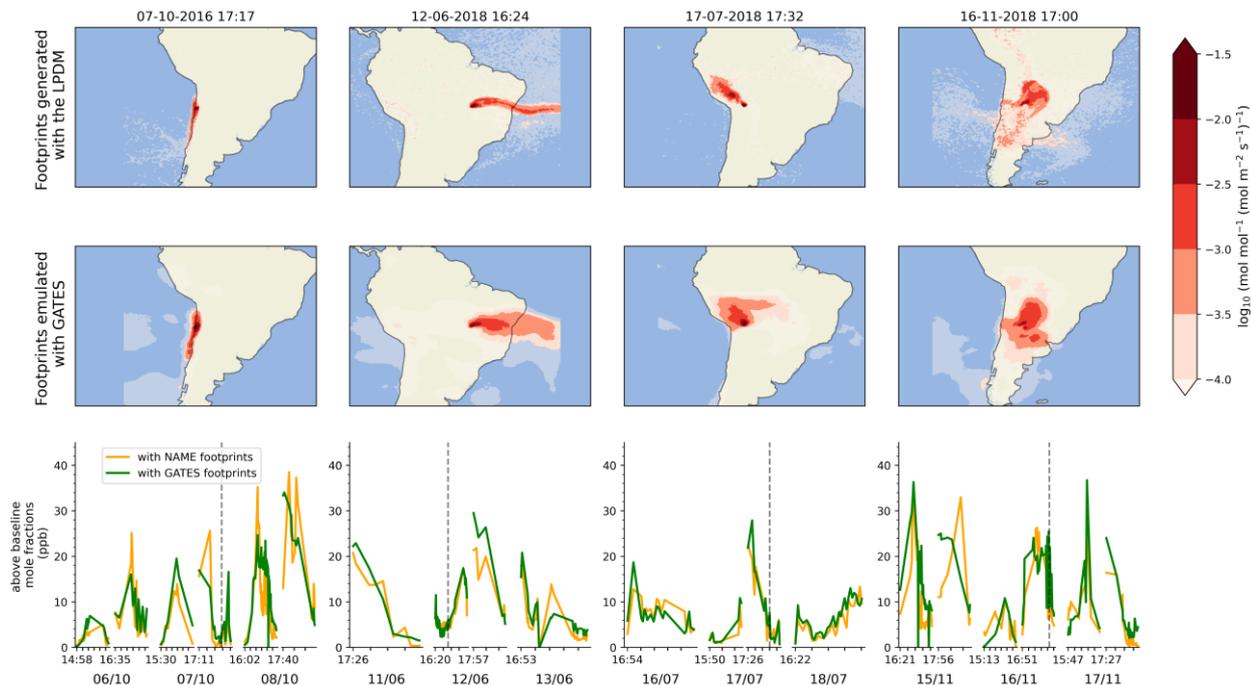
Figure 2: Example GATES outputs. Four LPDM-generated footprints from the test dataset (top row) and their GATES-emulated counterparts (second row). **The bottom row shows modelled above-baseline mole fractions,** derived with the LPDM footprints (orange) and the GATES footprints (green). Each box shows the timeseries for three days with breaks between days indicated by wider gaps between the lines. Each tick mark corresponds to two minute intervals. For days where the satellite scans the continent multiple times, we indicate the time of the first measurement of each overpass. The dashed line indicates the date and time of the footprints shown in the corresponding top two rows. Mole fractions are calculated by convolving each footprint with a prior, producing the estimated methane contribution in ppb from each grid cell to that particular measurement, and summing over the domain to calculate the total expected contribution from local sources (i.e. above the background baseline).

The definition of all the metrics used can be found in Appendix C. We evaluate the footprints themselves using the Correlation of Coefficient between the predicted and the true values, both in the original data space and in the logged space. The correlation in the original space captures how well the model reproduces the overall magnitude of the footprints, whereas applying the metric in the logged space gives more weight to low-magnitude values, often more diffuse and challenging to model. To quantify the spatial accuracy of the predicted footprints we use the metric Intersection over Union (IoU), which calculates the overlap between the predicted and true nonzero regions. The timeseries of simulated methane mole fractions is evaluated using standard metrics: the Coefficient of Correlation to assess the agreement between GATES and LPDM mole fractions; the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to quantify average and large errors, respectively; and the Mean Bias to indicate any systematic over- or underestimation in the predictions. The monthly emissions calculated with the inversion algorithm are evaluated using MAE and Mean Bias.

5.1 GATES-emulated footprints compared to the LPDM

Example emulated footprints (Fig. 2, middle row) and a set of goodness-of-fit metrics (Table 2) show that GATES captures well the overall shape and magnitude of the LPDM footprints over a range of meteorological regimes and locations in South America. However, the GATES outputs often tend to be somewhat smoother than the LPDM. This is primarily thought to be because the MSE loss function favours averaging in regions of high variability (similar smoothing effects have also been reported when using different architectures with MSE losses in LPDM emulation tasks (He et al., 2024). Moreover, since the LPDM is stochastic, the model outputs exhibit small-scale noise far from the measurement location, which is not reflected in

10

the emulated footprints. However, this noise has little impact on receptor mole fraction predictions. The model's connected architecture may also contribute to the smoothing effect, as similar behavior has been observed in weather applications using similar machine learning model framework (Lam et al., 2023).

350

We evaluate the spatial agreement of the emulated footprints using Intersection over Union (IoU) and the accuracy of the predictions using the correlation of coefficient (see Appendix C for all metrics), averaging both scores across all footprints in the test dataset. In the near-field (training region, S=50, Box A in Fig. 1), GATES achieves a mean IoU score (± one standard deviation) of 62% ± 23% and a mean correlation of coefficient of 0.70±0.13. Emulating the far field (full emulation domain,

355   S=200, Box B in Fig. 1) is a more difficult task, as the particles have travelled further from the measurement point and from each other. It is also a prediction area unseen to GATES during training. There is an expected drop in performance, particularly in the IoU as the LPDM footprints become noisier (IoU of 35% ± 12%, correlation coefficient to 0.68 ± 0.13). The drop in the correlation of coefficient is smaller as it is dominated by the large values very close to the measurement point.

360   We find that model performance is generally poorer in regions with more heterogeneous topography (see Fig. D1 in Appendix D), with footprints near the Andes scoring lower across all metrics. Chemical transport models are known to struggle over complex terrains (Brioude et al., 2012), potentially leading to more variance in the outputs and more difficult footprints to emulate. In addition, the seasonal and spatial distribution of observations in the training data might also influence performance: there are fewer training footprints in some of the mountainous areas, which could limit the learning. In contrast, the July-

365   August-September season, consistently achieves better metric scores, likely due to a higher number of available training samples, or potentially because the meteorology in this season is easier to capture (Fig D1).

Evaluation of footprints

Evaluation Size	Coefficient of correlation	Coefficient of correlation (logged)	IoU
$S = 50$	0.703	0.564	62.2%
$S = 200$	0.684	0.430	35.0%

Evaluation of mole fractions

Evaluation Size	Coefficient of correlation	MAE ppb (10^{-9})	RMSE ppb (10^{-9})	Mean Bias ppb (10^{-9})
$S = 50$	0.858	2.01	3.28	1.75
$S = 200$	0.808	2.89	4.40	0.724

Table 2. Evaluating GATES. Top: comparing footprints against the LPDM outputs. We show the mean score across footprints, averaged for the testing dataset, at the training region ($S = 50$) and the inference region ($S = 200$). Bottom: Evaluation of mole

370   **fraction timeseries, averaged across the testing dataset, also shown at the training and inference regions.**

5.2 Evaluating simulated mole fractions

In the inversion pipeline, each footprint is convolved with an emissions field (element-wise multiplication of both 2D fields and then summation over the whole area) to calculate the modelled atmospheric column-averaged mole fraction at the satellite observation location and time. Here, we convolve both sets of footprints (emulated and LPDM-derived) with the emissions

375   field from (Tunnicliffe et al., 2020), which includes anthropogenic (EDGAR (Emission Database for Global Atmospheric Research) v4.3.2 database (Janssens-Maenhout et al., 2019)), biomass burning (GFED (Global Fire Emissions Database) v4.1) and wetland (SWAMPS (Surface WAter Microwave Product Series) (Schroeder et al., 2015)).

Figure 2 (bottom panel) show examples of the timeseries emulated with GATES and the LPDM, for twelve days across the

380   testing dataset. The GATES-modelled measurements achieve a correlation coefficient of 0.81 with the LPDM-modelled data. The Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), of 2.89 ppb and 4.40 ppb respectively, are well below the estimated instrument GOSAT repeatability of 13 ppb (Parker et al., 2020). The errors between the two timeseries

are approximately normally distributed, with a small tendency for GATES to overpredict (Mean Bias 0.72 ppb). Figure E1 in Appendix E shows the modelled mole fractions in space and seasonally, revealing spatial patterns in the mean bias. The model over-predicts mole fractions in west Amazonia and in the South of Brazil, and underpredicts mole fractions over the North Andes, in Peru and Bolivia. These regions had already been identified as having higher errors the footprint-wise metrics (figure D1). 1.6% of the emulated mole fractions show an absolute error greater than 13 ppb. Figure F1 in Appendix F shows examples of instances where the mole fraction errors are large: These footprints are often near large cities, where the calculated mole fractions themselves are large and where the influence of emulation errors, particularly the above-mentioned smoothing, becomes more pronounced.

5.3 Inversion

Assessing the performance of the emulated footprints in a full methane flux inversion is key to ensuring that GATES is fit-for-purpose. Here, we follow a well-studied previous example, using GOSAT observations over and near Brazil (see (Tunnicliffe et al., 2020)) and an inversion set-up similar to (Western et al., 2021) and (Ganesan et al., 2014), who use a hierarchical Bayesian method to estimate monthly regional emissions. Brazil was chosen as a case study, because it is a large country where there are major and diverse CH_4 emissions from human and natural sources, which are not well understood (Tunnicliffe et al., 2020). We carry out this process for two inversion setups that are identical except for the footprints used: one uses the LPDM-generated footprints (those used in (Tunnicliffe et al., 2020)), the other uses the GATES-emulated footprints.

We use a hierarchical Bayesian Markov Chain Monte Carlo (HB-MCMC) inversion algorithm as described in Ganesan et al. (2014) and used for example by Western et al. (2021) and Ward et al. (2025). This method enables a better characterisation of uncertainty by applying "uncertainties in the uncertainties". We use the same parameters as Western et al. (2021) . In this set-up, the LPDM footprints quantify the contribution of local emissions (i.e. from within the domain) to the modelled methane measurement. The contributions from outside the domain ("background") must also be quantified. As well as the footprints, NAME outputs sensitivities for each of the boundaries of the domain, which get convolved with concentration "curtains" from the ECMWF CAMS reanalysis database (Inness et al., 2019) to calculate the background concentrations incoming from the edges of the domain. These contributions also get adjusted as part of the inversion. In this work, GATES emulates only the footprints – the boundary sensitivities are used as generated by NAME throughout, to isolate the effect of the emulator. We find that the yearly flux estimates calculated with GATES are in good agreement with the more computationally expensive LPDM-powered estimates: the inversion using the LPDM footprints estimates Brazil's mean yearly emissions of methane to be 41 ± 5 Tg yr^{-1} for 2016 and 49 ± 6 Tg yr^{-1} for 2018, and the inversion using the GATES footprints estimates 42 ± 5 Tg yr^{-1} and 53 ± 7 Tg yr^{-1}, respectively (see Table 3 for the disaggregated metrics). Figure 3(a) shows the estimated monthly emissions for Brazil across the two years, where the Mean Absolute Error (only for the test months) is 3.8 Tg yr^{-1} in 2016 and 6.3 Tg yr^{-1} in 2018, which is comparable to the a posteriori emission uncertainties. Any errors in the GATES emulation will be propagated to the inversion: the GATES-derived emissions are generally higher than the LPDM-derived ones, consistent with footprints being overall underpredicted in the model. There is also a difference in the magnitude of the uncertainties themselves, with the GATES-driven inversion presenting approximately 15% higher uncertainties in the mean yearly estimates. This additional uncertainty likely reflects the extra error due to the emulation, which is propagated through to the fluxes via our hierarchical Bayesian approach that explores model-data uncertainty as part of the inversion.

Figure 3(b) shows the difference between the derived emissions (posterior) and the initial estimate (prior) for 2018 for the two inversions, and Fig. 3(c) shows the difference between these adjustments. The fluxes are generally in good spatial agreement (Mean Absolute Error in the derived fluxes of 0.003 μmol m^{-2} s^{-1} compared to mean value of 0.010 μmol m^{-2} s^{-1} across the two years). Both inversions most notably increased the prior estimate in Brazil's east coast and in the Pantanal region on the border with Bolivia. The GATES posterior shows larger emissions coming from Western Amazonia than the LPDM-driven posterior. This region consistently shows higher errors in the footprint-wise metrics (figure D1) and consistent underprediction of the mole fractions (figure E1), consistent with deriving higher emissions. Simulated methane concentrations based on the LPDM-derived posterior and the GATES-derived posterior both show a significantly improved fit to the observed GOSAT XCH4

12

compared to the prior-based simulations (see Appendix G). The two posteriors show consistent corrections to the prior, with the GATES-based simulations showing only a small positive mean bias of 2.69 ppb relative to those from LPDM.

Overall, our flux estimates for Brazil are consistent within uncertainties with several previous studies that have used GOSAT and inversions based on LPDMs or Eulerian models (Janardanan et al., 2020; Wilson et al., 2021). Whilst within uncertainties, the mean GATES and LPDM estimates are somewhat higher than that of Tunnicliffe et al. (2020). This difference occurs, despite the same satellite observations and LPDM because Tunnicliffe et al. (2020) also incorporated surface data from Ragged Point, Barbados, and used a different inverse method.

Year	Mean yearly emissions ($Tg\ yr^{-1}$)		MAE (Tg)	Mean Bias (Tg)
	With LPDM footprints	With GATES footprints	*For the monthly timeseries*	
2016 (full year)	41.4 ± 4.6	42.4 ± 5	3.17	0.94
2016 (test months only)	38.0 ± 4	39.3 ± 4.6	3.84	1.35
2018	49.3 ± 6	52.7 ± 7.0	6.26	1.44
2018 (excluding Dec)	46.5 ± 5	46.7 ± 5.7	5.47	0.22

Table 3 – Mean yearly methane emissions for Brazil in 2016 and 2018, estimated using the LPDM footprints and GATES emulated footprints, and mean errors between the two monthly timeseries. We show that the Mean Absolute Error for the monthly timeseries is comparable to the uncertainty in the yearly estimates. Results are displayed for the full 2016 (Jan-Dec), the 2016 test set (April-Dec), the full 2018, and 2018 excluding Dec. December 2018 has <100 valid observations, and therefore the inferred emissions have very high uncertainties which is reflected in the annual mean. This month was also excluded from the analysis conducted by Tunnicliffe et al. (2020).
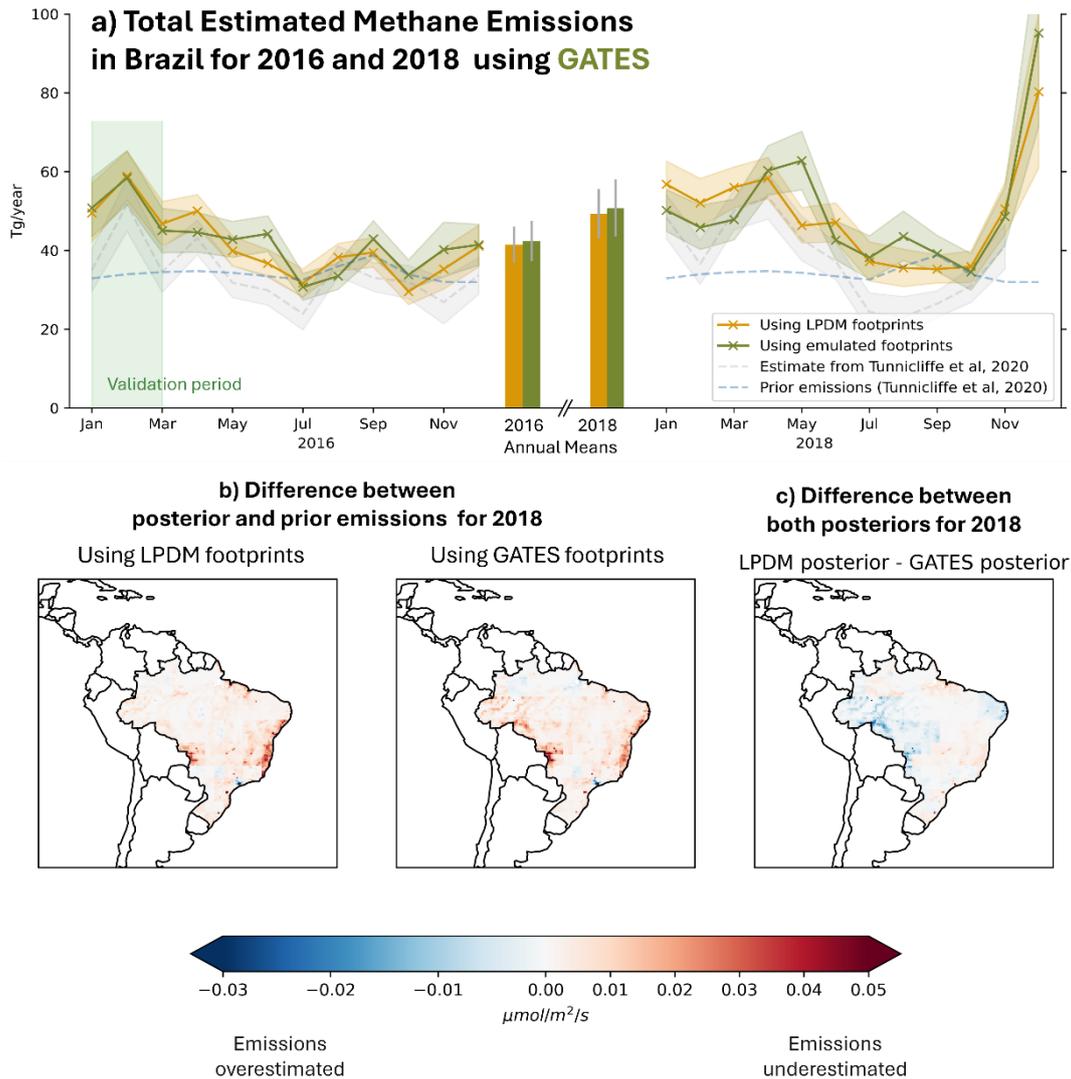
Fig. 3. Deriving methane emissions for Brazil, in time (top panel) and space (bottom panels). (a) Estimated methane emissions, for 2016 and 2018. We compare the emissions estimates derived in an inversion using the LPDM footprints (yellow) with those emulated by GATES (green). The emissions inferred with the emulated footprints are well within the posterior inversion uncertainty of the emissions inferred with the traditional method. Note that we report the annual means (bars at the center of the plot) for the full 2016 (i.e. including the validation set January-March) to allow for like-for-like comparisons with 2018. We also show the monthly priors used by (Tunnicliffe et al., 2020) and their derived emissions. (b) Difference between the inversion posterior and prior fluxes for 2018, using the LPDM footprints (left) and the GATES footprints (right). The difference highlights the areas that were underestimated or overestimated in the initial estimate. The fluxes calculated with the GATES footprints (right) show good spatial agreement with those calculated from the traditional method (left). (c) Difference between the two posteriors for 2018. Blue shows the areas where the GATES posterior estimates higher emissions than the LPDM posterior, and red shows the opposite.

14

6 Discussion and Conclusions

We present GATES, a novel ML-driven approach to emulating LPDM footprints using graph neural networks. We demonstrate its use as a stand-alone application, but most importantly within the primary intended workflow, to infer GHG emissions using
460   satellite mole fraction observations. The high consistency of the estimates of Brazil's methane emissions from the ML-enabled pipeline with those based on the physical model, at a fraction of the computational cost (>1000 times faster), demonstrate the potential of this emulator to accelerate greenhouse gas estimations.

One of the main limitations of GATES, as identified for similar GNNs in weather applications (Lam et al., 2023), is that the
465   outputs tend to be smoother than the LPDM footprints. This is likely due to the use of an MSE-based loss function, which encourages less variance in the outputs rather than accurately reproducing fine-scale spatial details, particularly in areas of higher uncertainty. This effect can be observed in Fig. 2, especially in the second footprint, where the predicted plume is in the correct region and direction, but is broader than that of the LPDM. Smoother footprints may lead to higher inference errors in regions of more point-like sources, such as cities or oil and gas infrastructure (see Fig. S6).
470

Alternative architectures may help to improve future versions of GATES. GATES is a deterministic model, generating a single footprint for a set of inputs, whereas LPDMs are inherently probabilistic. This difference contributes to the smoothing effect, as GATES tends to predict an average outcome rather than generating multiple plausible outputs. Incorporating probability into the emulation would likely produce sharper outputs and contribute to a more rigorous quantification of uncertainty.
475   Understanding uncertainty would also improve explainability, highlighting the areas where the model presents more variance. Further improvements to the architecture of the model could also increase its predicting capabilities, e.g. by integrating multi-meshes like (Lam et al., 2023) to capture dispersion at different scales, or adaptive mesh refinement (Pfaff et al., 2020) to encourage higher resolution predictions only around the footprint, or in complex geographies.

480   Emulating the boundary sensitivities is a key step to fully integrate GATES into an inversion pipeline. These boundary sensitivities are used to calculate the background concentrations, incoming from outside the domain. In this work, we use the boundary sensitivities generated by the LPDM throughout, isolating the effect of the column-averaged footprint emulator. Future work will address this limitation, developing a boundary condition emulator.

485   Here, we show GATES in application, using GOSAT data to infer emissions from Brazil, but we expect that it could be readily extended to other regions and instruments. As there are only minimal differences in the setup of LPDM simulations for retrievals of similar short-wave infrared satellites, GATES should also be easily re-trainable to several existing and upcoming mapping satellites (e.g., MethaneSat, TROPOMI, CO2M). The growing volume of space-based GHG measurements offers vast opportunities for improving emissions estimates. As these datasets continue to expand, ML approaches like GATES
490   provide a promising avenue to leverage this data more efficiently, enhancing global efforts in tracking emissions at increasing temporal and spatial scales and with reduced latency.

Appendix A

## Distribution of the data in space and time - Training Data
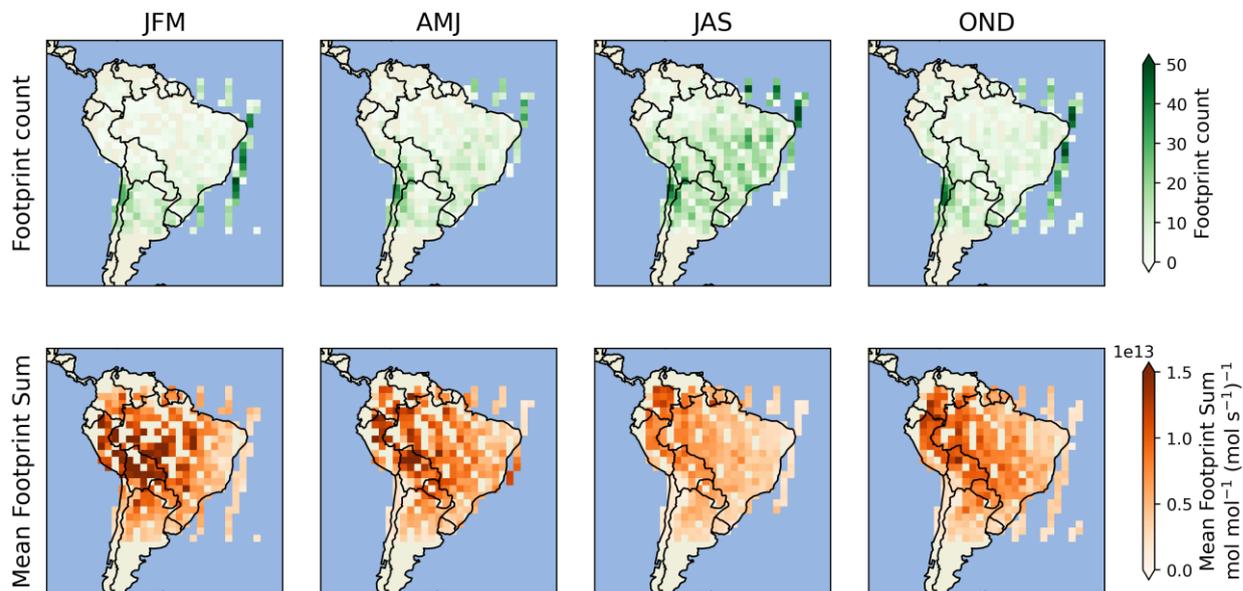
Distribution of the data in space and time - Testing data
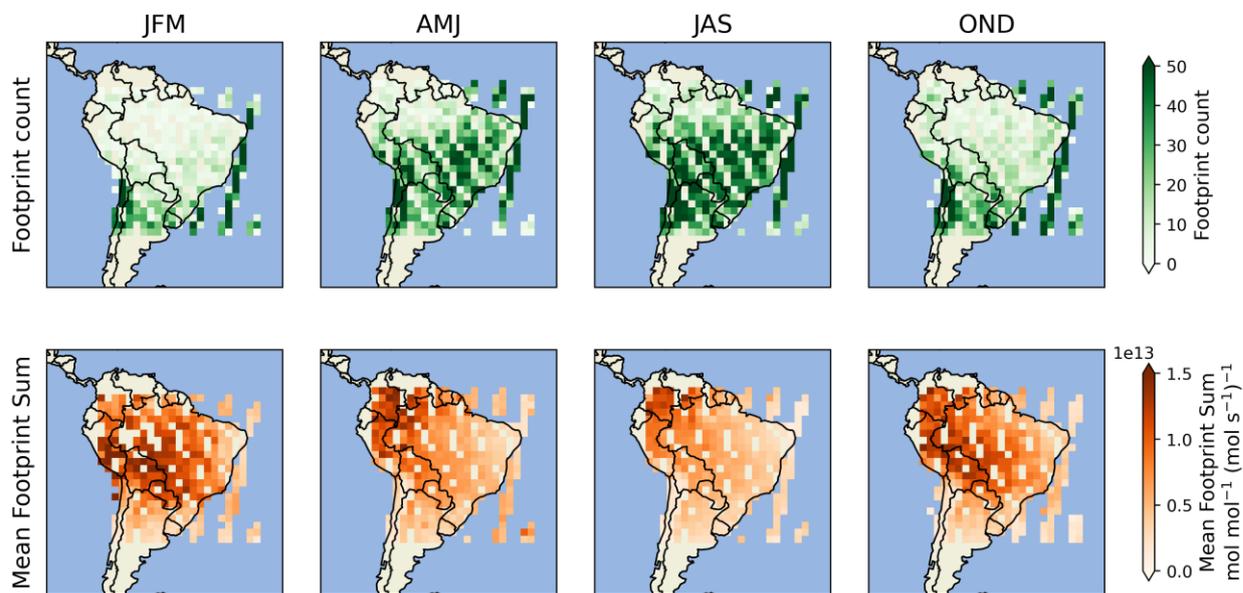
500 **Fig. A1. Distribution in space and by season (see key below) of the footprint datasets, split in training (2014-2015, top panel) and validation and testing (2016 and 2018, bottom panel).** For each dataset, the count of samples in each 2° bin (**top subpanel**, green) shows seasonal variations in the data available, with more coverage in July-August-September, as well as spatial biases, with more observations in the southern regions and over the Atlantic. Note that although we train with footprints over the ocean, in the inversion only land observations are used, and we report metrics for those. **The bottom subpanels** show
505 the mean sum of the footprints in that bin, over the spatial domain (the sensitivity in each grid cell is weighted by the area). Higher mean footprint sums indicate less particle dispersion, often due to slower winds or geographical barriers. For example, mean sums are higher near the Andes, where the high topography prevents the particles from advecting. (Key: JFM = January, February, March, AMJ = April, May, June, JAS = June, July, August, OND = October, November, December)

Appendix B
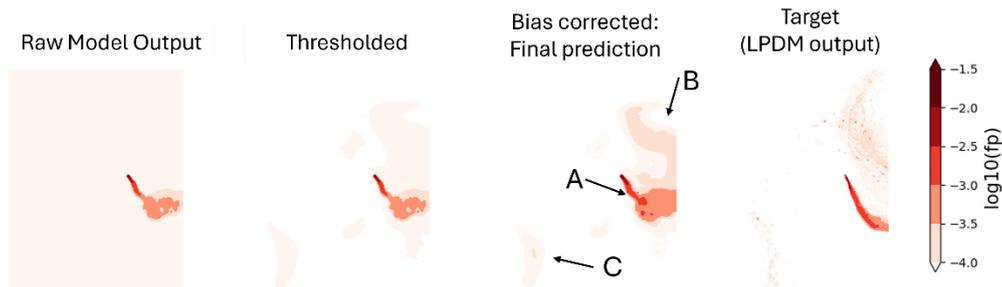
Fig. B1. Example of the post-processing method in GATES. The raw prediction from GATES for a particular set of inputs from the test set is shown in the left-most panel. The target footprint predicted by NAME is shown in the right-most panel for comparison. Thresholding is first applied to the emulated footprint, removing all background values under threshold. This threshold is decided from the validation data. A distribution-wise bias correction method, calibrated with the validation data, is then applied, which corrects the magnitude of the footprint values particularly on the tail of the footprint (A) and noisy regions (B and C).
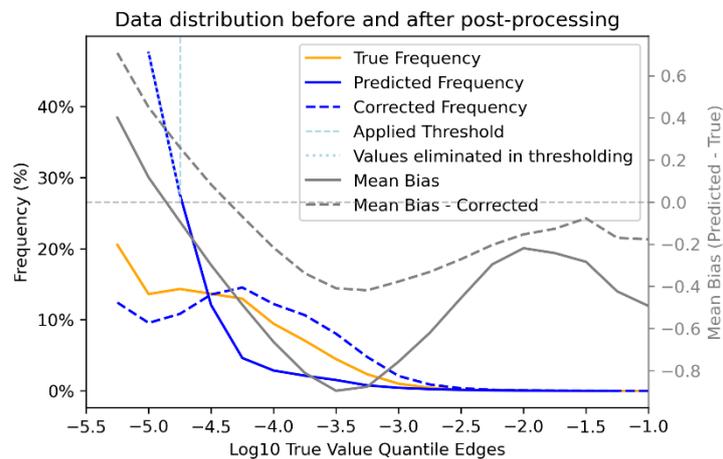
Fig. B2. Impact of the bias correction on the data distribution of the test set. In orange we show the logged distribution of the test data (wherever it is non-zero) for April-December 2016 (left axis), split into 20 bins. The distribution of the values predicted directly by GATES is shown with the blue line: there is a very high proportion of small background values. The solid gray line, on the right axes, indicates the mean bias of the raw GATES predictions in each quantile of the true data (in the logged space). GATES tends to under-predict particularly the mid-range values ($\log(f) \approx$ -3.5. The post-processing steps are then applied. First we remove background values under a threshold tuned with the validation set (1.757×10^{-5}), indicated by the dashed light blue line. Bias correction is then applied on the remaining values, with the resulting distribution shown as the dashed blue line. The corrected presents an improved fit against the true frequency, and the mean bias of the two distribution with a dashed grey line on the right axes, shows that the mid- and high-range values are better captured.

The raw prediction from GATES for a particular set of inputs from the test set is shown in the left-most panel. The target footprint predicted by NAME is shown in the right-most panel for comparison. Thresholding is first applied to the emulated

18

footprint, removing all background values under threshold. This threshold is decided from the validation data. A distribution-wise bias correction method, calibrated with the validation data, is also applied, which corrects the magnitude of the footprint values particularly on the tail of the footprint (A) and noisy regions (B and C).

535

Appendix C

The emulated footprints are evaluated at three levels: comparing the emulated footprints to the LPDM-generated ones, evaluating the simulated methane column mole fractions, and within an inversion. We use the metrics below to evaluate the model at different levels, where b_i is the true value and $\hat{b}_i$ is the emulated value.

1. Pearson's Correlation Coefficient:

$$r(\boldsymbol{b}, \hat{\boldsymbol{b}}) = \frac{\sum_{i=1}^{n}(b_i - \bar{b})(\hat{b}_i - \bar{\hat{b}})}{\sqrt{\sum_{i=1}^{n}(b_i - \bar{b})^2}\sqrt{\sum_{i=1}^{n}(\hat{b}_i - \bar{\hat{b}})^2}}$$

The correlation of coefficient is applied to:
- footprint-to-footprint evaluation. The coefficient of correlation is calculated for each footprint so that $\hat{\boldsymbol{b}}$ represents the flattened values and comparing them against the flattened corresponding LPDM footprint. The footprint-wise correlations of coefficient can then be shown as averages across different locations or time-periods (section 5.1). It can be applied in the native dataspace, or in the logged dataspace $\left(r\left(\log(\boldsymbol{b}), log(\hat{\boldsymbol{b}})\right)\right)$, wherever the footprint is non-zero. This metric ignores the shape of the predicted footprints and evaluates the magnitude only.
- Mole fraction evaluation: The coefficient of correlation is calculated for the whole dataset of predicted mole fractions. It ignores any time correlation between the data-points.

2. Intersection Over Union (IoU), also known as Figure of Merit in Space or Jaccard index: quantifies the spatial agreement between the emulated and the true footprints by computing the ratio of the overlapping area to the total area covered by either. It is applied on a binarised version of $\boldsymbol{b}$ above a particular threshold $\tau=0$, so that $B_{i,j}=1$ if $\boldsymbol{b}_{i,j}>0$, or $B_{i,j}=0$ otherwise. IoU score ranges from 0 to 1 (or 0% to 100%), where 1 indicates perfect overlap and 0 indicates no overlap at all. The IoU is calculated for each footprint, and averaged across samples in section 5.1.

$$\text{IoU}(B, \hat{B}) = \frac{|B \cap \hat{B}|}{|B \cup \hat{B}|}$$

3. Mean Absolute Error: measures the average absolute difference between the predicted and true values. It is used to evaluate the above-baseline mole fractions (section 5.2), and also the monthly methane emissions for Brazil (section 5.3).

$$MAE(\boldsymbol{b}, \hat{\boldsymbol{b}}) = \frac{1}{n}\sum_{i=1}^{n}|b_i - \hat{b}_i|$$

4. Mean Squared Error and Root Mean Squared Error: measure the average of the squared differences between the predicted and true values. The Root MSE (RMSE) is the square root of MSE, the error in the same units as the original data. expressing It is used to evaluate the above-baseline mole fractions (section 5.2).

$$MSE(\boldsymbol{b}, \hat{\boldsymbol{b}}) = \frac{1}{n}\sum_{i=1}^{n}(b_i - \hat{b}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(b_i - \hat{b}_i)^2}$$

5. Mean Bias: measures the average difference between predicted and true values, indicating whether the model systematically overestimates or underestimates with respect to the true value. It is used to evaluate the above-baseline mole fractions (section 5.2), and also the monthly methane emissions for Brazil (section 5.3).

$$\text{MBE}(\boldsymbol{b}, \hat{\boldsymbol{b}}) = \frac{1}{n} \sum_{i=1}^{n} (\hat{b}_i - b_i)$$
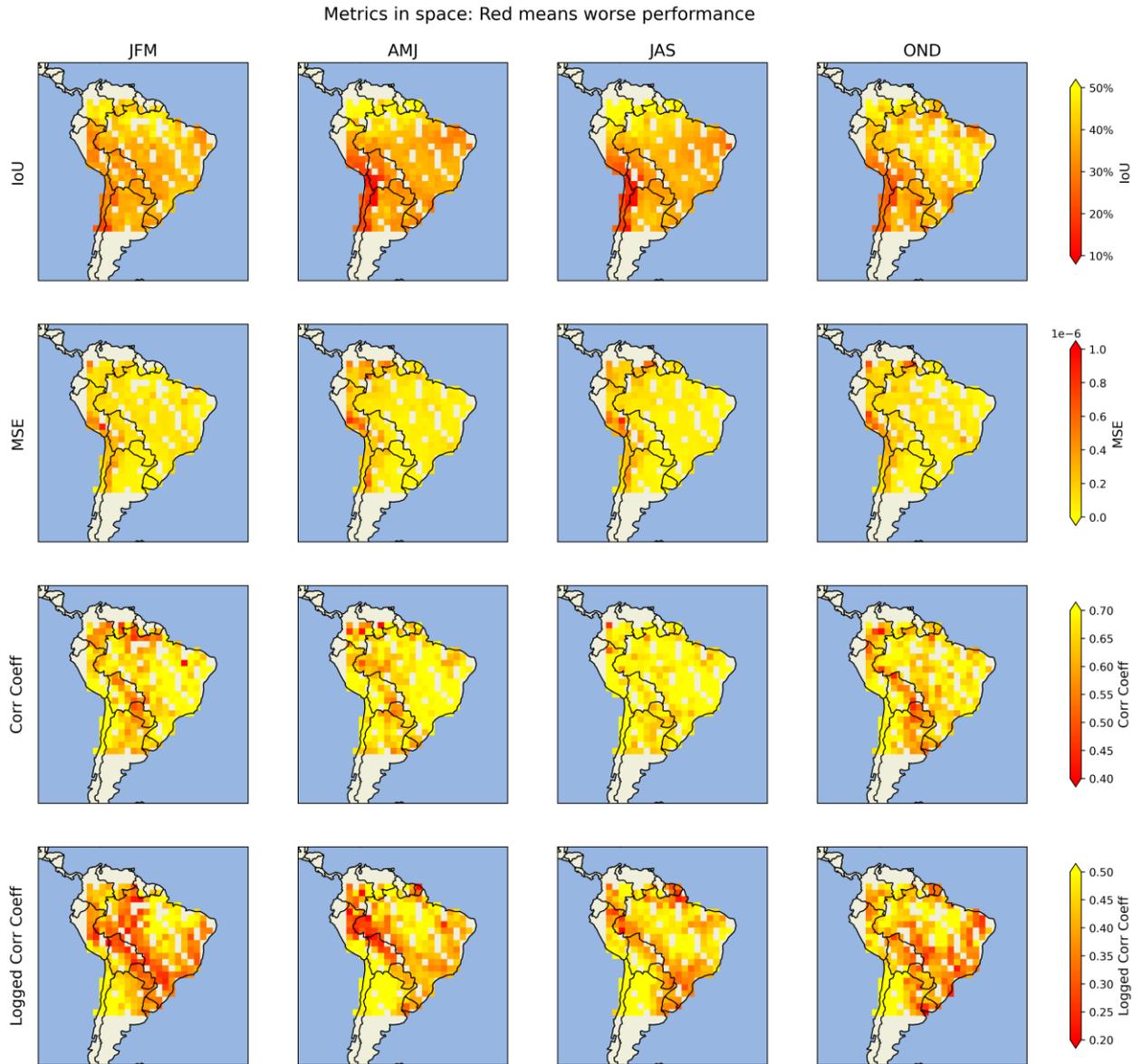
Fig. D1. – Mean footprint metric scores disaggregated in space and time. Red indicates worse performance for that metric. The footprints are binned by season (see key below) and by the coordinates of the satellite measurements in 2° bins, showing here the mean metrics for each bin in 2016 and 2018. (Key: JFM = January, February, March, AMJ = April, May, June, JAS = June, July, August, OND = October, November, December)

580

585
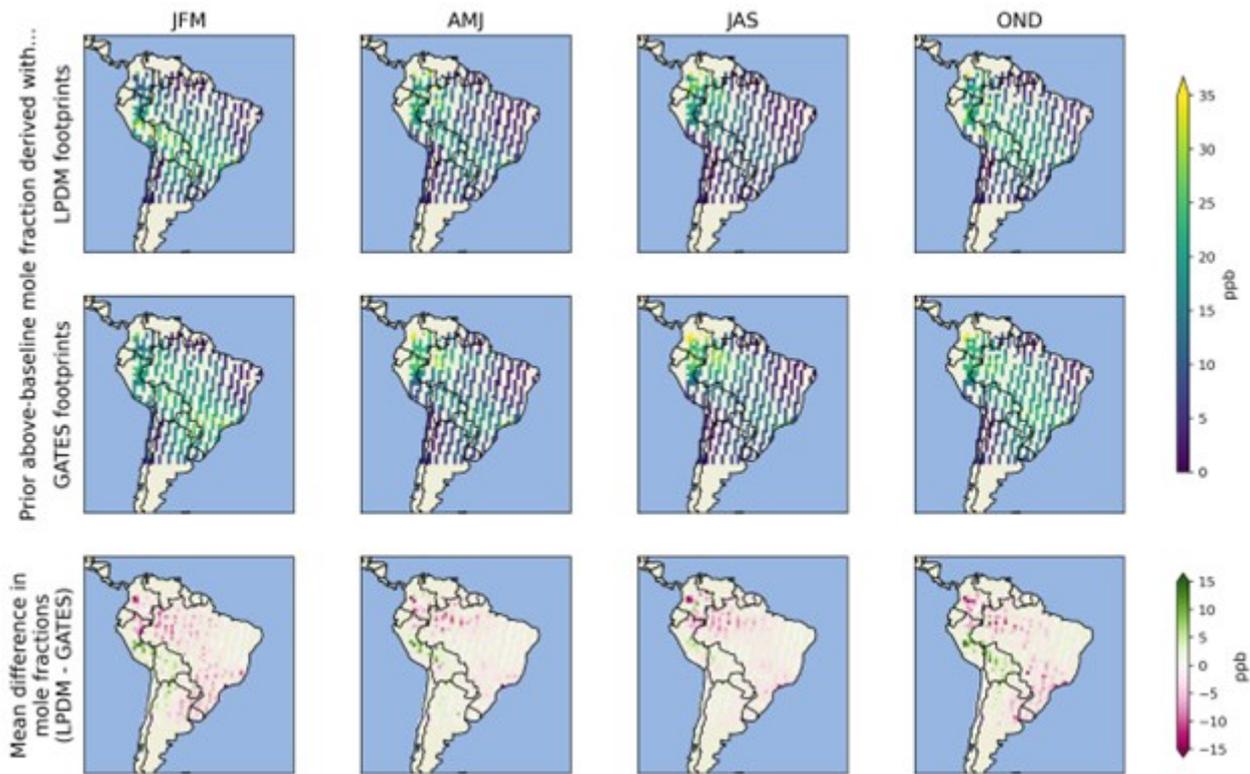
Fig. E1. – Seasonal mean above-baseline mole fractions, derived with the LPDM footprints (top row), with GATES footprints (middle row) and the difference between the two (bottom row). The mole fractions are derived by convolving a monthly emissions prior with each footprint (see section 5.2 or Tunnicliffe et al. (2020) for a description of the prior). Here, the mole fractions are binned by season (see key below) and in 1° bins, averaged for both 2016 and 2018. The mole fractions calculated with GATES footprints (middle row) show similar spatial patterns to those calculated with LPDM footprints (top row). The two are compared in the bottom row, where positive values (green) indicate mole fractions that were underpredicted by GATES, and negative values (pink) indicate mole fractions that were overpredicted by GATES.
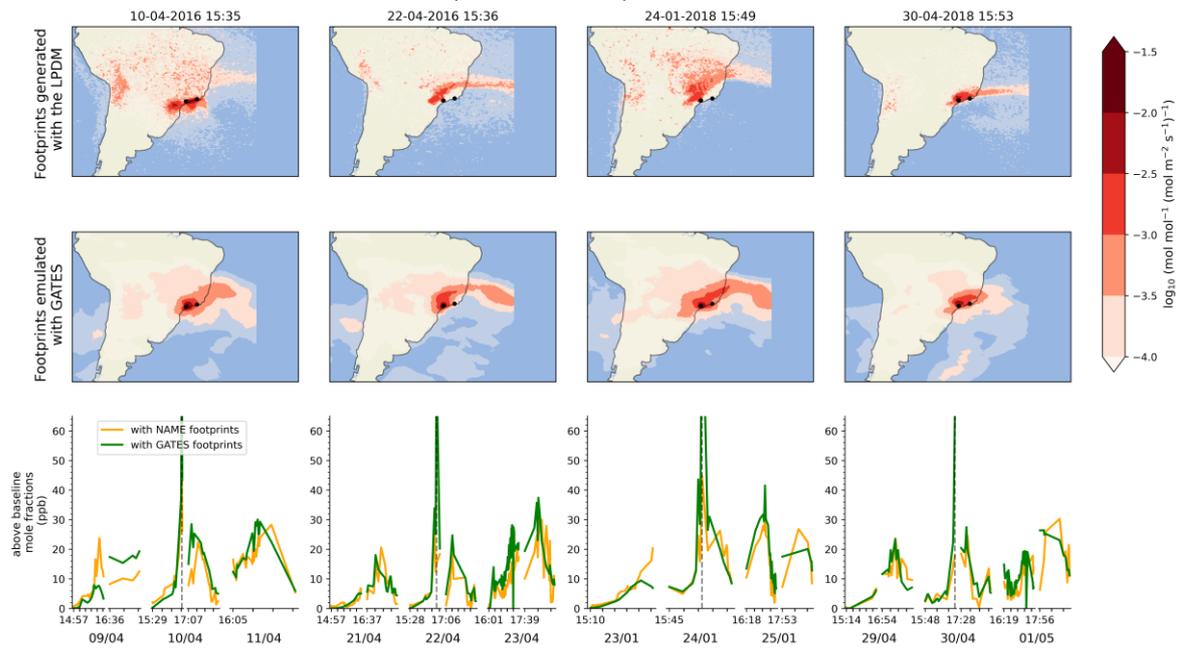
(Key: JFM = January, February, March, AMJ = April, May, June, JAS = June, July, August, OND = October, November, December)

Four footprints that underpredict the mole fractions
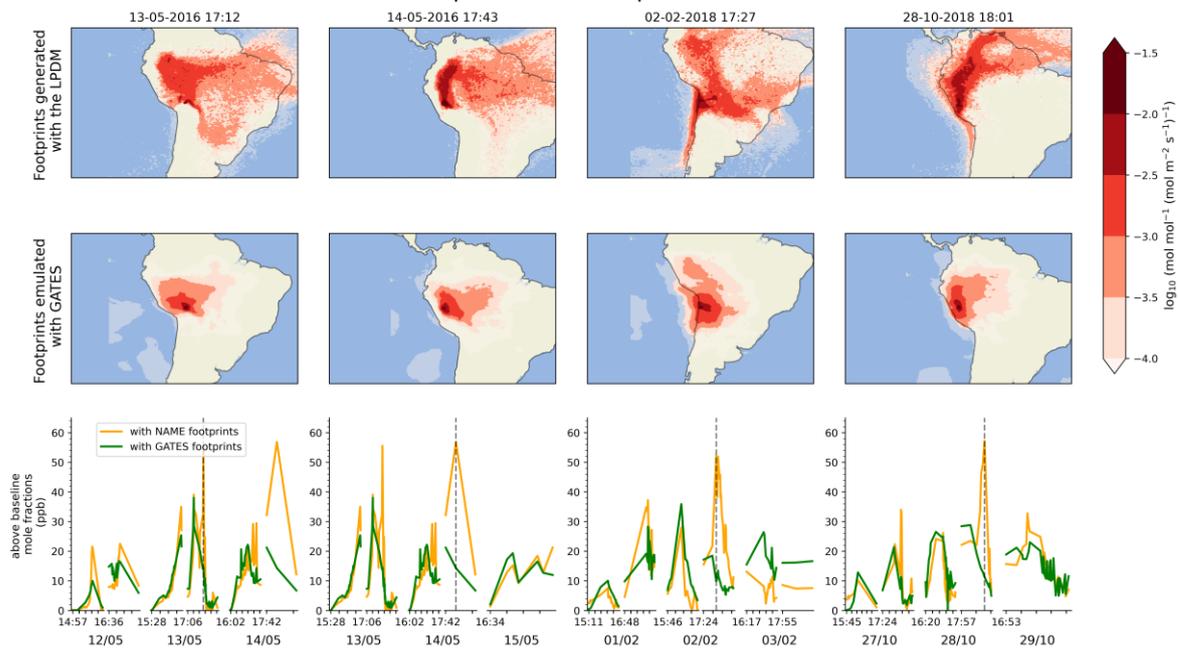
Four footprints that overpredict the mole fractions

600

Fig. F1. Two sets of footprints that lead to the largest errors in the emulated mole fractions, through underprediction (top panel) or overprediction (bottom panel). In each panel, we show the LPDM footprint (top row), the GATES-emulated footprint (bottom row) and the derived mole-fraction timeseries, for the day of the footprint, the day before and the day after. **Top panel:** Footprints for measurements near the central Andes area accumulate higher errors than other areas (see figure D1), and for these four footprints GATES fails to emulate complex dispersion patterns from difficult meteorology and topography. The LDPM footprints capture dispersion over the Amazon basin, where estimated emissions are large, leading to a big difference between the two estimates. **Bottom panel**: The smoothing effect is particularly visible here, as sharp, narrow plumes in the LPDM footprints over or near Sao Paulo and Rio de Janeiro (black dots) are emulated to be wider and with softer edges. When convolved with the prior, the large emissions from the cities inflate the small errors in the GATES footprints, leading to overpredictions.

Appendix G:

615 The emissions derived from the inversion can be compared against the actual observations taken by GOSAT. In Fig. G1 (left panel), we show the distribution of prior modelled mole fractions (i.e. footprints × prior fluxes + prior background) for 2018, calculated with the GATES footprints and the LPDM footprints. We also show the distribution of GOSAT XCH4 observations, and the posterior modelled mole fractions (i.e. footprints × posterior fluxes + posterior background) derived with either set of footprints. Both posteriors fit the observations significantly better than the priors, with most of the improvement attributed to

620 an increase in the background mole fractions. (Tunnicliffe et al., 2020) and (Thompson et al., 2025) also identify that adjusting the boundary together with the fluxes is key to obtaining well-fitting posteriors. The posteriors derived with either set of footprints are very similar, and show similar error distributions when compared to the observations (Fig. G1, right panel). The prior derived with GATES has a small bias to overpredict mole fractions, which translates into a small bias in the posteriors.
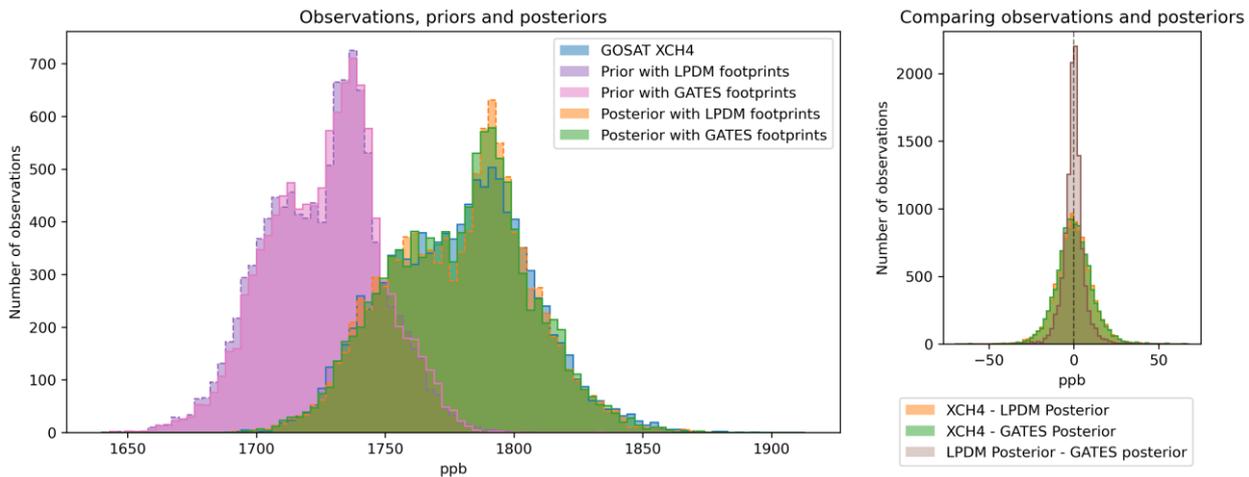
625

Fig. G1. Comparing the distributions of the GOSAT observations, prior and posterior mole fractions in 2018. Left panel: Distributions for a priori modelled methane mole fractions (purple and pink), GOSAT XCH4 observations (blue) and the posterior mole fractions (orange and green). We show the prior and posterior pair generated with the LPDM
630 footprints (purple and orange respectively, dashed edges) and the posterior and prior pair generated with the GATES footprints (purple and green respectively, solid edges). Both posteriors fit the observations well. **Right panel: Distribution of errors between the XCH4 observations and the mole fractions calculated from both sets of posteriors (in orange and green), and differences between the mole fractions derived with both posteriors (brown).** Both posteriors show similar error patterns between the estimated emissions and the actual XCH4 observations. The errors between the two posteriors (brown)
635 are normally distributed, with a small positive mean bias (2.69 ppb) which indicates a small overpredicting tendency in GATES.

Code and data availability

The code used to implement, train and evaluate GATES is available at DOI https://zenodo.org/doi/10.5281/zenodo.16679175 (Fillola et al., 2025), and future updates and documentation will be made available as a free access repository at https://github.com/elenafillo/GATES_LPDM_emulator. The trained model for South America described above can also be found in the repository. The full dataset of LPDM footprints generated with NAME can be found at DOI https://zenodo.org/doi/10.5281/zenodo.16748754 (Fillola and Tunnicliffe, 2025). The UM meteorology in its original format is publicly available, hosted by the UK Centre for Environmental Data Analysis (CEDA) Archive, from July 2015 at https://data.ceda.ac.uk/badc/name_nwp/data/global (Met Office, 2025a, b) under a UK Open Government Licence 3.0. Earlier dates are available on request from the Met Office. The meteorology can be extracted from its original format with the code at https://github.com/elenafillo/extract_iris_met (Fillola and Clark, 2025). The software used for the inversion can be found at DOI https://doi.org/10.5281/zenodo.6834888 (Rigby et al., 2022). The GOSAT XCH4 retrievals are located at https://dx.doi.org/10.5285/18ef8247f52a4cb6a14013f8235cc1eb (Parker et al., 2020). Users can train GATES by extracting the NAME meteorology and the provided footprints, or by replacing these with other meteorological datasets and LPDM model outputs.

Author contribution

Conceptualization: EF, MR, RSR

Methodology: EF, MR, RSR

Investigation: EF

Software: EF

Data curation: EF, RT

Visualization: EF

Supervision: MR, RSR

Writing – original draft: EF

Writing – review & editing: EF, RSR, RT, JNC, NK, AG, MR

Competing interests

The authors declare that they have no conflict of interest.

References

680 Alexe, M., Bergamaschi, P., Segers, A., Detmers, R., Butz, A., Hasekamp, O., Guerlet, S., Parker, R., Boesch, H., Frankenberg, C., Scheepmaker, R. A., Dlugokencky, E., Sweeney, C., Wofsy, S. C., and Kort, E. A.: Inverse modelling of CH_4 emissions for 2010–2011 using different satellite retrieval products from GOSAT and SCIAMACHY, Atmospheric Chemistry and Physics, 15, 113–133, https://doi.org/10.5194/acp-15-113-2015, 2015.

Ba, J. L., Jamie, R. K., and Hinton, G. E.: Layer Normalization, https://doi.org/arXiv:1607.06450v1, 2016.

685 Baker, D. F.: TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO2 fluxes, Global Biogeochemical Cycles, 20, 1988–2003, https://doi.org/10.1029/2004GB002439, 2006a.

Baker, D. F.: Variational data assimilation for atmospheric CO2, Tellus B: Chemical and Physical Meteorology, 58, 359–365, https://doi.org/10.1111/j.1600-0889.2006.00218.x, 2006b.

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V. F., Malinowski, M., Tacchetti, A., Raposo, 690 D., Santoro, A., Faulkner, R., Gülçehre, Ç., Song, H. F., Ballard, A. J., Gilmer, J., Dahl, G. E., Vaswani, A., Allen, K. R., Nash, C., Langston, V., Dyer, C., Heess, N., Wierstra, D., Kohli, P., Botvinick, M. M., Vinyals, O., Li, Y., and Pascanu, R.: Relational inductive biases, deep learning, and graph networks, CoRR, abs/1806.01261, https://doi.org/arXiv:1806.01261, 2018.

Bergamaschi, P., Karstens, U., Manning, A. J., Saunois, M., Tsuruta, A., Berchet, A., Vermeulen, A. T., Arnold, T., Janssens-695 Maenhout, G., Hammer, S., Levin, I., Schmidt, M., Ramonet, M., Lopez, M., Lavric, J., Aalto, T., Chen, H., Feist, D. G., Gerbig, C., Haszpra, L., Hermansen, O., Manca, G., Moncrieff, J., Meinhardt, F., Necki, J., Galkowski, M., O'Doherty, S., Paramonova, N., Scheeren, H. A., Steinbacher, M., and Dlugokencky, E.: Inverse modelling of European CH_4 emissions during 2006–2012 using different inverse models and reassessed atmospheric observations, Atmospheric Chemistry and Physics, 18, 901–920, https://doi.org/10.5194/acp-18-901-2018, 2018.

700 Bousquet, P. P.: Regional Changes in Carbon Dioxide Fluxes of Land and Oceans Since 1980, Science, 290, 1342–1346, https://doi.org/10.1126/science.290.5495.1342, 2000.

Brioude, J., Angevine, W. M., McKeen, S. A., and Hsie, E.-Y.: Numerical uncertainty at mesoscale in a Lagrangian model in complex terrain, Geoscientific Model Development, 5, 1127–1136, https://doi.org/10.5194/gmd-5-1127-2012, 2012.

Cartwright, L., Zammit-Mangion, A., and Deutscher, N. M.: Emulation of greenhouse-gas sensitivities using variational 705 autoencoders, https://doi.org/arXiv:2112.12524v1, 2021.

Dadheech, N., He, T.-L., and Turner, A. J.: [preprint] High-resolution greenhouse gas flux inversions using a machine learning surrogate model for atmospheric transport, https://doi.org/10.5194/egusphere-2024-2918, 2024.

Essery, R., Best, M., and Cox, P.: MOSES 2.2 Technical Documentation, Hadley Centre, Met Office (https://jules.jchmr.org/sites/default/files/2023-06/JULES-HCTN-30.pdf), 2001.

710 European Commission and United States of America: Global Methane Pledge, (https://www.ccacoalition.org/resources/global-methane-pledge), 2021.

Fasoli, B., Lin, J. C., Bowling, D. R., Mitchell, L., and Mendoza, D.: Simulating atmospheric tracer concentrations for spatially distributed receptors: Updates to the Stochastic Time-Inverted Lagrangian Transport model's R interface (STILT-R version 2), Geoscientific Model Development, 11, 2813–2824, https://doi.org/10.5194/gmd-11-2813-2018, 2018.

715 Feng, L., Palmer, P. I., Parker, R. J., Lunt, M. F., and Bösch, H.: Methane emissions are predominantly responsible for record-breaking atmospheric methane growth rates in 2020 and 2021, Atmospheric Chemistry and Physics, 23, 4863–4880, https://doi.org/10.5194/acp-23-4863-2023, 2023.

Fillola, E. and Clark, J.: extract_iris_met, , https://doi.org/10.5281/ZENODO.16634862, 2025.

Fillola, E. and Tunnicliffe, R.: LPDM footprint dataset - Brazil, https://zenodo.org/doi/10.5281/zenodo.16748754, 2025.

720 Fillola, E., Santos-Rodriguez, R., Manning, A., O'Doherty, S., and Rigby, M.: A machine learning emulator for Lagrangian particle dispersion model footprints: a case study using NAME, Geoscientific Model Development, 16, 1997–2009, https://doi.org/10.5194/gmd-16-1997-2023, 2023.

Fillola, E., Clark, J., and Keshtmand, N.: GATES: A Graph-Neural-Network Atmospheric Transport Emulation System, , https://doi.org/10.5281/ZENODO.16679175, 2025.

725 Ganesan, A. L., Rigby, M., Zammit-Mangion, A., Manning, A. J., Prinn, R. G., Fraser, P. J., Harth, C. M., Kim, K.-R., Krummel, P. B., Li, S., Mühle, J., O'Doherty, S. J., Park, S., Salameh, P. K., Steele, L. P., and Weiss, R. F.: Characterization of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods, Atmospheric Chemistry and Physics, 14, 3855–3864, https://doi.org/10.5194/acp-14-3855-2014, 2014.

Ganesan, A. L., Rigby, M., Lunt, M. F., Parker, R. J., Boesch, H., Goulding, N., Umezawa, T., Zahn, A., Chatterjee, A., Prinn, R. G., Tiwari, Y. K., van der Schoot, M., and Krummel, P. B.: Atmospheric observations show accurate reporting and little growth in India's methane emissions, Nature Communications, 8, 836, https://doi.org/10.1038/s41467-017-00994-7, 2017.

He, T.-L., Dadheech, N., Thompson, T. M., and Turner, A. J.: FootNet v1.0: Development of a machine learning emulator of atmospheric transport, earthArXiv [preprint], https://doi.org/10.31223/X5197G, 2024.

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmospheric Chemistry and Physics, 19, 3515–3556, https://doi.org/10.5194/acp-19-3515-2019, 2019.

Jacob, D. J., Varon, D. J., Cusworth, D. H., Dennison, P. E., Frankenberg, C., Gautam, R., Guanter, L., Kelley, J., McKeever, J., Ott, L. E., Poulter, B., Qu, Z., Thorpe, A. K., Worden, J. R., and Duren, R. M.: Quantifying methane emissions from the global scale down to point sources using satellite observations of atmospheric methane, Atmospheric Chemistry and Physics, 22, 9617–9646, https://doi.org/10.5194/acp-22-9617-2022, 2022.

Janardanan, R., Maksyutov, S., Tsuruta, A., Wang, F., Tiwari, Y. K., Valsala, V., Ito, A., Yoshida, Y., Kaiser, J. W., Janssens-Maenhout, G., Arshinov, M., Sasakawa, M., Tohjima, Y., Worthy, D. E. J., Dlugokencky, E. J., Ramonet, M., Arduini, J., Lavric, J. V., Piacentino, S., Krummel, P. B., Langenfelds, R. L., Mammarella, I., and Matsunaga, T.: Country-Scale Analysis of Methane Emissions with a High-Resolution Inverse Model Using GOSAT and Surface Observations, Remote Sensing, 12, https://doi.org/10.3390/rs12030375, 2020.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., Bergamaschi, P., Pagliari, V., Olivier, J. G. J., Peters, J. A. H. W., van Aardenne, J. A., Monni, S., Doering, U., Petrescu, A. M. R., Solazzo, E., and Oreggioni, G. D.: EDGAR v4.3.2 Global Atlas of the three major greenhouse gas emissions for the period 1970–2012, Earth System Science Data, 11, 959–1002, https://doi.org/10.5194/essd-11-959-2019, 2019.

Joint CEOS-CGMS Working Group on Climate - Greenhouse Gas Task Team.: Roadmap for a Coordinated Implementation of Carbon Dioxide and Methane Monitoring from Space, CEOS-CGMS

(https://ceos.org/document_management/Publications/Publications-and-Key-Documents/Atmosphere/CEOS_CGMS_GHG_Roadmap_Issue_2_V1.0_FINAL.pdf), 2024.

755 Jones, A., Thomson, D., Hort, M., and Devenish, B.: The U.K. Met Office's next-generation atmospheric dispersion model, NAME III, Air Pollution Modeling and Its Application XVII, 580–589, https://doi.org/10.1007/978-0-387-68854-1_62, 2007.

Kaminski, T., Heimann, M., and Giering, R.: A coarse grid three-dimensional global inverse model of the atmospheric transport: 1. Adjoint model and Jacobian matrix, Journal of Geophysical Research: Atmospheres, 104, 18535–18553, https://doi.org/10.1029/1999jd900147, 1999.

760 Keisler, R.: Forecasting Global Weather with Graph Neural Networks, ArXiv [preprint], https://doi.org/arXiv:2202.07575, 2022.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P.: Learning skillful medium-range global weather forecasting, Science, 382, 1416–1421, https://doi.org/10.1126/science.adi2336, 2023.

765 Leip, A., Skiba, U., Vermeulen, A., and Thompson, R. L.: A complete rethink is needed on how greenhouse gas emissions are quantified for national reporting, Atmospheric Environment, 174, 237–240, https://doi.org/10.1016/j.atmosenv.2017.12.006, 2018.

Loshchilov, I. and Hutter, F.: Decoupled Weight Decay Regularization, in: International Conference on Learning Representations, https://doi.org/arXiv:1711.05101, 2019.

770 Maasakkers, J. D., Jacob, D. J., Sulprizio, M. P., Scarpelli, T. R., Nesser, H., Sheng, J.-X., Zhang, Y., Hersher, M., Bloom, A. A., Bowman, K. W., Worden, J. R., Janssens-Maenhout, G., and Parker, R. J.: Global distribution of methane emissions, emission trends, and OH concentrations and trends inferred from an inversion of GOSAT satellite data for 2010–2015, Atmospheric Chemistry and Physics, 19, 7859–7881, https://doi.org/10.5194/acp-19-7859-2019, 2019.

Met Office: UKCP18 Guidance: Bias correction, , 775 (https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18-guidance---how-to-bias-correct.pdf), 2018.

Met Office: Global NWP meteorological data for Met Office NAME dispersion model (Mk9: July 2015 - 2017) (Mk9), CEDA (Centre for Environmental Data Analysis) (https://data.ceda.ac.uk/badc/name_nwp/data/global/UMG_Mk9) [dataset], https://catalogue.ceda.ac.uk/uuid/2bb4d76ed2fa4fc2af3fbbca6eb80965/, 2025a.

780 Met Office: Global NWP meteorological data for Met Office NAME dispersion model (Mk10: June 2017 - May 2022) (Mk10), CEDA (Centre for Environmental Data Analysis) (https://data.ceda.ac.uk/badc/name_nwp/data/global/UMG_Mk10) [dataset], https://catalogue.ceda.ac.uk/uuid/7d0fff9f59b94a3da347e3ae10bd8fc1/, 2025b.

Nair, V., Hinton, E., and, G.: Rectified linear units improve restricted boltzmann machines, in: ICML'10: Proceedings of the 27th International Conference on International Conference on Machine Learning, 807–814, 785 https://doi.org/10.5555/3104322.3104425, 2010.

Parker, R. J., Webb, A., Boesch, H., Somkuti, P., Barrio Guillo, R., Di Noia, A., Kalaitzi, N., Anand, J. S., Bergamaschi, P., Chevallier, F., Palmer, P. I., Feng, L., Deutscher, N. M., Feist, D. G., Griffith, D. W. T., Hase, F., Kivi, R., Morino, I., Notholt, J., Oh, Y.-S., Ohyama, H., Petri, C., Pollard, D. F., Roehl, C., Sha, M. K., Shiomi, K., Strong, K., Sussmann, R., Té, Y., Velazco, V. A., Warneke, T., Wennberg, P. O., and Wunch, D.: A decade of GOSAT Proxy satellite \chemCH_4 observations, 790 Earth System Science Data, 12, 3383–3412, https://doi.org/10.5194/essd-12-3383-2020, 2020.

Peters, W., Miller, J. B., Whitaker, J., Denning, A. S., Hirsch, A., Krol, M. C., Zupanski, D., Bruhwiler, L., and Tans, P. P.: An ensemble data assimilation system to estimate CO2 surface fluxes from atmospheric trace gas observations, Journal of Geophysical Research, 110, https://doi.org/10.1029/2005jd006157, 2005.

Pfaff, T., Fortunato, M., Sanchez-Gonzalez, A., and Battaglia, P.: Learning Mesh-Based Simulation with Graph Networks, in: International Conference on Learning Representations, https://doi.org/arXiv:2010.03409, 2020.

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, Geoscientific Model Development, 12, 4955–4997, https://doi.org/10.5194/gmd-12-4955-2019, 2019.

Rigby, M., Tunnicliffe, R., lukewestern, hanchawn, ag12733, aliceramsden, Jones, G., Young, D., Ward, R., Angharad, ANickless-Bristol, and joe-pitt: ACRG-Bristol/acrg: ACRG v0.2.0 (v0.2.0), , https://doi.org/10.5281/zenodo.6834888, 2022.

Roten, D., Wu, D., Fasoli, B., Oda, T., and Lin, J. C.: An interpolation method to reduce the computational time in the stochastic Lagrangian particle dispersion modeling of spatially dense XCO2 retrievals, Earth and Space Science, 8, https://doi.org/10.1029/2020ea001343, 2021.

Sanchez-Gonzalez, A., Godwin, J., Pfaff, T., Ying, R., Leskovec, J., and Battaglia, P. W.: Learning to simulate complex physics with graph networks, in: Proceedings of the 37th International Conference on Machine Learning, 8459–8468, https://arxiv.org/abs/2002.09405, 2020.

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.

Saunois, M., Martinez, A., Poulter, B., Zhang, Z., Raymond, P. A., Regnier, P., Canadell, J. G., Jackson, R. B., Patra, P. K., Bousquet, P., Ciais, P., Dlugokencky, E. J., Lan, X., Allen, G. H., Bastviken, D., Beerling, D. J., Belikov, D. A., Blake, D. R., Castaldi, S., Crippa, M., Deemer, B. R., Dennison, F., Etiope, G., Gedney, N., Höglund-Isaksson, L., Holgerson, M. A., Hopcroft, P. O., Hugelius, G., Ito, A., Jain, A. K., Janardanan, R., Johnson, M. S., Kleinen, T., Krummel, P. B., Lauerwald, R., Li, T., Liu, X., McDonald, K. C., Melton, J. R., Mühle, J., Müller, J., Murguia-Flores, F., Niwa, Y., Noce, S., Pan, S., Parker, R. J., Peng, C., Ramonet, M., Riley, W. J., Rocher-Ros, G., Rosentreter, J. A., Sasakawa, M., Segers, A., Smith, S. J., Stanley, E. H., Thanwerdas, J., Tian, H., Tsuruta, A., Tubiello, F. N., Weber, T. S., van der Werf, G. R., Worthy, D. E. J., Xi, Y., Yoshida, Y., Zhang, W., Zheng, B., Zhu, Q., Zhu, Q., and Zhuang, Q.: Global Methane Budget 2000–2020, Earth System Science Data, 17, 1873–1958, https://doi.org/10.5194/essd-17-1873-2025, 2025.

Scarpelli, T. R., Jacob, D. J., Grossman, S., Lu, X., Qu, Z., Sulprizio, M. P., Zhang, Y., Reuland, F., Gordon, D., and Worden, J. R.: Updated Global Fuel Exploitation Inventory (GFEI) for methane emissions from the oil, gas, and coal sectors: evaluation with inversions of atmospheric methane observations, Atmospheric Chemistry and Physics, 22, 3235–3249, https://doi.org/10.5194/acp-22-3235-2022, 2022.

Schmidli, J., Frei, C., and Vidale, P. L.: Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods, International Journal of Climatology, 26, 679–689, https://doi.org/10.1002/joc.1287, 2006.

835 Schroeder, R., McDonald, K. C., Chapman, B. D., Jensen, K., Podest, E., Tessler, Z. D., Bohn, T. J., and Zimmermann, R.: Development and Evaluation of a Multi-Year Fractional Surface Water Data Set Derived from Active/Passive Microwave Remote Sensing Data, Remote Sensing, 7, 16688–16732, https://doi.org/10.3390/rs71215843, 2015.

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., and Ngan, F.: NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System, Bulletin of the American Meteorological Society, 96, 2059–2077,
840 https://doi.org/10.1175/BAMS-D-14-00110.1, 2015.

Thompson, R. L., Krishnankutty, N., Pisso, I., Schneider, P., Stebel, K., Sasakawa, M., Stohl, A., and Platt, S.: Efficient use of a Lagrangian Particle Dispersion Model for atmospheric inversions using satellite observations of column mixing ratios, EGUsphere, 2025, 1–28, https://doi.org/10.5194/egusphere-2025-147, 2025.

Tunnicliffe, R. L., Ganesan, A. L., Parker, R. J., Boesch, H., Gedney, N., Poulter, B., Zhang, Z., Lavrič, J. V., Walter, D.,
845 Rigby, M., Henne, S., Young, D., and O'Doherty, S.: Quantifying sources of Brazil's CH4 emissions between 2010 and 2018 from Satellite Data, Atmospheric Chemistry and Physics, 20, 13041–13067, https://doi.org/10.5194/acp-20-13041-2020, 2020.

Uber Technologies Inc: H3: Uber's Hexagonal Hierarchical Spatial Index (3.7.7), , PyPi (https://pypi.org/project/h3/3.7.7/), 2024.

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., Vries, J. de, Otter, G., Claas, J., Eskes, H. J., Haan, J. F. de, Kleipool, Q.,
850 Weele, M. van, Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: 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, https://doi.org/10.1016/j.rse.2011.09.027, 2012.

Vojta, M., Plach, A., Thompson, R. L., and Stohl, A.: A comprehensive evaluation of the use of Lagrangian particle dispersion
855 models for inverse modeling of greenhouse gas emissions, Geoscientific Model Development, 15, 8295–8323, https://doi.org/10.5194/gmd-15-8295-2022, 2022.

Ward, R. H., Western, L. M., Tunnicliffe, R. L., Fillola, E., Tsuruta, A., Aalto, T., and Ganesan, A. L.: Uncertainty and retrieval sensitivity in TROPOMI-based methane inversions over the North Slope of Alaska, EGUsphere, 2025, 1–38, https://doi.org/10.5194/egusphere-2025-4230, 2025.

860 Western, L. M., Ramsden, A. E., Ganesan, A. L., Boesch, H., Parker, R. J., Scarpelli, T. R., Tunnicliffe, R. L., and Rigby, M.: Estimates of North African Methane Emissions from 2010 to 2017 Using GOSAT Observations, Environ. Sci. Technol. Lett., 8, 626–632, https://doi.org/10.1021/acs.estlett.1c00327, 2021.

Wilson, C., Chipperfield, M. P., Gloor, M., Parker, R. J., Boesch, H., McNorton, J., Gatti, L. V., Miller, J. B., Basso, L. S., and Monks, S. A.: Large and increasing methane emissions from eastern Amazonia derived from satellite data, 2010–2018,
865 Atmospheric Chemistry and Physics, 21, 10643–10669, https://doi.org/10.5194/acp-21-10643-2021, 2021.

Worden, J. R., Pandey, S., Zhang, Y., Cusworth, D. H., Qu, Z., Bloom, A. A., Ma, S., Maasakkers, J. D., Byrne, B., Duren, R., Crisp, D., Gordon, D., and Jacob, D. J.: Verifying Methane Inventories and Trends With Atmospheric Methane Data, AGU Advances, 4, e2023AV000871, https://doi.org/10.1029/2023AV000871, 2023.