https://doi.org/10.5194/egusphere-2026-3361
Preprint. Discussion started: 28 September 2026

GATES-Background 0.2.0: Emulating background greenhouse gas mole fractions for regional atmospheric inverse modelling with graph neural networks

Nawid Keshtmand[1,2], Elena Fillola[1,2], Jeff Clark[2], Raul Santos-Rodriguez[2], and Matthew Rigby[1]

[1]Atmospheric Chemistry Research Group, School of Chemistry, University of Bristol, Bristol, UK
[2]School of Engineering Mathematics, University of Bristol, Bristol, UK

Correspondence: Nawid Keshtmand (nawid.keshtmand@bristol.ac.uk) and Elena Fillola (elena.fillolamayoral@bristol.ac.uk)

Abstract. Regional atmospheric trace gas inverse modelling frameworks depend on accurately simulating mole fractions, which result from transporting surface fluxes within a domain and carrying mole fractions into the region from its boundaries. With the aim of improving the computational efficiency of inverse modelling pipelines, recent machine learning approaches have successfully emulated Lagrangian Particle Dispersion Model (LPDM) footprints (source-receptor relationships) at substantially reduced computational cost, compared to the numerical model simulations. However, the estimation of the background component of the simulated mole fractions, an important contributor to overall mole fraction variability, has received comparatively little attention. In this study, we build on the GATES model (Graph-Neural-Network Atmospheric Transport Emulation System, Fillola et al. (2026b)), which emulates LPDM outputs, and present GATES-Background 0.2.0, a complementary emulator for background mole fraction contributions to observations across a regional domain driven by meteorological data and large-scale background estimates from a global methane reanalysis.

GATES-Background is trained and evaluated to emulate background concentrations for GOSAT observations over South America and demonstrates strong agreement with reference calculations generated from the propagation of boundary conditions to the measurement location. The emulator generally captures daily variability, seasonal structure, and regional gradients in background mole fraction contributions, with errors that are relatively small and spatially localized, compared to the variability in the simulated background values (root-mean-square errors of 5.3–7.4 ppb, depending on season, compared to GOSAT XCH4 retrieval error on the order of 13 ppb and estimated variability due to boundary conditions within this region of approximately 15 - 20 ppb). The proposed approach reduces computational cost by several orders of magnitude relative to physics-based background calculations, enabling scalable and near–real-time application. When combined with GATES or other LPDM footprint emulators, this work will enable efficient machine learning–based approximation of both components of the forward model used in regional LPDM-based greenhouse gas inversion systems.

https://doi.org/10.5194/egusphere-2026-3361
Preprint. Discussion started: 28 September 2026

1 Introduction

National or regional emissions estimates can be calculated using "bottom-up" inventories constructed from activity data and process models (e.g. Saunois et al., 2025). Complementing these approaches, top-down methods use atmospheric observations to constrain emissions through inverse modelling. These systems depend on atmospheric transport models to calculate the sensitivity of measured mole fractions to surface fluxes. Lagrangian Particle Dispersion Models (LPDMs) are widely used for this purpose in regional inversion frameworks (e.g. Ganesan et al., 2017; Thompson and Stohl, 2014; Manning et al., 2011; Hu et al., 2017). By simulating the backward-in-time transport of air parcels from an observation point, LPDMs can produce spatial sensitivity maps, commonly referred to as footprints, which quantify the influence of upwind surface fluxes on observed mole fractions (Manning et al., 2011). In addition to surface footprints, LPDMs can generate sensitivity functions to mole fractions at spatial domain boundaries (e.g. Lunt et al., 2016), or from the distribution of particles in the atmosphere at some point in time prior to a measurement (e.g. Thompson and Stohl, 2014). These boundary condition sensitivities represent the contribution of mole fractions to an observation, excluding the influence of near-field fluxes. Background mole fractions at each measurement point can be calculated by multiplying these sensitivities by mole fraction estimates at the the spatial or temporal boundary (e.g., provided by some global model into which data have been assimilated). Together, surface and boundary condition sensitivities provide a total mole fraction estimate based on:

$$y_{\mathrm{total}} = H_{\mathrm{flux}}\, x_{\mathrm{flux}} + H_{\mathrm{bc}}\, x_{\mathrm{bc}} \tag{1}$$

Here, y_{total} denotes the mole fraction at the measurement location (e.g., from a satellite or in situ). The operator H_{flux} corresponds to the LPDM-derived surface footprint, which quantifies the sensitivity of the observation to emissions within the domain, and x_{flux} represents the corresponding surface flux. The second term on the right-hand side accounts for the large-scale background contribution. H_{bc} denotes the boundary sensitivity operator, while x_{bc} represents the prescribed boundary mole fraction field, typically obtained from global chemical transport model-derived products such as those provided through the Copernicus Atmosphere Monitoring Service (CAMS) (Inness et al., 2019). This formulation explicitly separates the total observed mole fraction into contributions from local surface fluxes and the large-scale background, allowing both terms to be estimated in an inversion.

Throughout this paper, mole fractions will refer to weighted column means for the GOSAT satellite using pressure weights and averaging kernels from Parker et al. (2020). For methane, this quantity is typically referred to as XCH_4. However, the proposed methodology could similarly be applied to mole fractions for other parts of the atmosphere (e.g., for surface observations), or to other gases, provided an appropriate training dataset were used.

The method outlined in this paper has been applied to LPDM simulations of spatial boundary conditions, in which the sensitivity of observations within a domain to mole fractions on a "curtain" surrounding it are estimated. However, it could readily be applied to temporal boundary conditions using the same methodology. Some regional inverse modelling systems rely on estimating background mole fractions from the observations themselves, but this approach is not considered here (see, for example, Manning et al., 2011; Gerrand et al., 2025).

https://doi.org/10.5194/egusphere-2026-3361
Preprint. Discussion started: 28 September 2026

LPDM simulations are computationally expensive for large observational datasets, with a single footprint calculation typically requiring on the order of 10 CPU-minutes. Therefore, the rapidly increasing volume of satellite-based greenhouse gas observations presents a significant computational challenge for inversion systems relying on direct LPDM simulations (Veefkind et al., 2012). To address this limitation, machine learning (ML) methods are increasingly being explored as emulators of computationally expensive physics-based simulations (e.g., Fillola et al., 2023; He et al., 2023; Fillola et al., 2026b).

In atmospheric transport modelling, ML-based emulators must map high-dimensional meteorological inputs to high-dimensional spatial outputs, while preserving spatial structure and physically meaningful relationships between inputs and outputs. Graph Neural Networks (GNNs) are well suited to this task, as they can naturally operate on structured grids and explicitly represent local interactions through message passing (Battaglia et al., 2018; Bacciu et al., 2020). Recent work by Fillola et al. (2026b) demonstrated the feasibility of GNN-based LPDM emulation through the GATES model (Graph-Neural-Network Atmospheric Transport Emulation System), showing that atmospheric footprints can be predicted from meteorological and topographical inputs at a computational cost several orders of magnitude lower than that of the underlying dispersion model. However, Fillola et al. (2026b) relied on boundary condition sensitivities calculated using the numerical model.

Here, we develop a complementary model, GATES-Background, which emulates background mole fractions that influence observations across a regional domain. The model predicts the background contribution to trace gas mole fractions using meteorological and atmospheric composition state information as inputs. We demonstrate the approach over the region of South America for XCH4 methane observations, showing that the emulator reproduces key temporal and spatial characteristics of the background mole fraction, including the main temporal and spatial gradients and seasonality. When combined with existing GNN-based LPDM footprint emulators, like GATES, this work enables efficient emulation of both the local and background components of the forward model used in top-down greenhouse gas inversion systems, removing the need to run a physics-based model beyond generating training datasets.

2 Method

Architecture We use the same encoder-processor-decoder architecture as Fillola et al. (2026b), which is similar to architectures used in meteorological forecasting applications (Keisler, 2022; Lam et al., 2022; Pfaff et al., 2020). A visual overview of the architecture and additional details are provided in Fillola et al. (2026b). The inputs, including meteorological and topographical data, are arranged in a latitude-longitude grid that represents the native data space. In the encoder, this data is encoded into a triangular, lower-resolution mesh, and in the processor each mesh node is updated iteratively using a GNN block (Battaglia et al., 2018), which aggregates information from neighbouring nodes.

The main difference between our work and Fillola et al. (2026b) lies in the decoder architecture and prediction objective. While the prior work focused on predicting a full spatial footprint at the original grid resolution, our work predicts a single aggregated value, specifically the background mole fraction at each observation point($H_{\mathrm{bc}}\, x_{\mathrm{bc}}$). To achieve this, we convert the data from the mesh into a regular grid by using a distance-weighted mean between the values of the three closet mesh nodes to each grid node, and replace the node decoder with a convolutional decoder consisting of two convolutional layers. Each layer

https://doi.org/10.5194/egusphere-2026-3361
Preprint. Discussion started: 28 September 2026

uses a kernel size of 3, a stride of 2, and padding of 1, followed by a ReLU (Rectified Linear Unit) activation, which is an activation function that sets all negative values to zero while keeping positive values unchanged. Convolutional layers are well suited to spatially structured data such as geophysical fields, as they exploit local spatial correlations by applying shared filters across the grid. A kernel size of 3 allows each output value to incorporate information from a local 3×3 neighbourhood, while a stride of 2 progressively reduces spatial resolution, summarizing information over larger areas. Padding ensures consistent treatment of grid boundaries. Together, these design choices enable the model to aggregate spatial information efficiently while preserving key spatial patterns. The convolutional decoder therefore provides a computationally efficient means of capturing spatial relationships and producing a compact representation suitable for predicting aggregated quantities, while retaining information relevant to the underlying spatial structure.

Data Background methane mole fractions are derived by convolving the LPDM-generated boundary condition sensitivities with global, observation-constrained mole fraction fields from CAMS (Inness et al., 2019). The boundary sensitivities are generated by the UK Met Office Lagrangian Particle Dispersion Model Numerical Atmospheric dispersion Modelling Environment (NAME Jones et al., 2007) for Greenhouse gases Observing SATellite (GOSAT Hamazaki et al., 2005) observations over South America during 2014–2017 (Tunnicliffe et al., 2020; Parker et al., 2020). The north and south boundary curtains consist of 20×190 grid cells, spanning 20 vertical levels from 500 m to 19,500 m at 1,000 m intervals, while the east and west curtains comprise 20×357 grid cells at the same vertical resolution. The CAMS methane mole fractions are interpolated to match the horizontal and vertical resolution of the particle location arrays along each boundary.

The ML model inputs include meteorological variables at 7 atmospheric levels, as well as topographic and coordinate information, following Fillola et al. (2026b). For computational efficiency, inputs are cropped to a 50×50 grid-cell domain (approximately 1000×1500 km) centered on the satellite measurement location.

To provide the ML model with information on the regional background mole fraction, the CAMS boundary field is subsampled. Values are provided to the model at four altitudes (1100, 1700, 2400, and 3200 m) at the horizontal midpoint of each boundary. We apply a detrending procedure by subtracting the average southern CAMS field at 1.5 km above the surface from the summed background mole fraction, thereby isolating in-domain variability from the long-term trend. We train on data for 2014-2016 ($\sim$18000 data points after sampling by using every third GOSAT footprint to thin the highly correlated dataset), and test on data for 2017 ($\sim$21000 data points after sampling). The model predicts the detrended background mole fraction at each GOSAT measurement point onto which the southern CAMS mole fraction is re-added to provide an estimate of the full background mole fraction.

Training details The model is trained for 300 epochs with a batch size of 5, using the Adam optimizer with a learning rate of 1×10^{-6}. We train the model using the mean squared error between the predicted mole fraction and the true mol fraction. We train and run the model on a 32GB NVIDIA V100 GPU which takes approximately 8 hours.

Evaluation metrics To assess the performance of our model, we examine the root mean squared error (RMSE), bias, and the random error (σ) given by Equations 2 to 4, and the relationship between the three quantities can be seen in Equation 5:

https://doi.org/10.5194/egusphere-2026-3361
Preprint. Discussion started: 28 September 2026

$$\mathrm{RMSE}(\hat{y}_i, y_i) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{2}$$

$$\mathrm{Bias}(\hat{y}_i, y_i) = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i) \tag{3}$$

$$\sigma_{\mathrm{error}}(\hat{y}_i, y_i) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[(\hat{y}_i - y_i) - \mathrm{Bias}]^2} \tag{4}$$

$$\mathrm{RMSE}^2 = \mathrm{Bias}^2 + \sigma_{\mathrm{error}}^2 \tag{5}$$

3 Results and discussion

We analyse the performance of our model for different seasons of the year, defined here for convenience as January-February-March (JFM), April-May-June (AMJ), July-August-September (JAS), and October-November-December (OND). The magnitude of the background emulator error can be contextualized by comparison with both instrumental uncertainty and "true" (NAME-predicted) background variability. Typical uncertainty in GOSAT XCH_4 retrievals is approximately 13 ppb (Parker et al., 2020). Within our domain, the standard deviation of the background contribution to XCH_4 estimated by NAME is approximately 15 - 20 ppb, depending on the season. Most of this variability is due to the latitudinal gradient as the satellite traverses the domain (Figure 1).

Table 1 summarizes the seasonal RMSE, bias, random error (σ_{error}) and standard deviation of the predicted background methane mole fractions, as well as the standard deviation of the background mole fractions obtained from NAME. RMSE values range from 5.3 to 7.4 ppb and are dominated by the random error component, as reflected by the close agreement between RMSE and σ_{error}. Across all seasons, model biases remain within ± 3 ppb, with the overall aggregated annual bias calculated at -1.1 ppb. The highest performance occurs during JAS (July–September), which exhibits both the lowest bias (0.1 ppb) and the smallest random error (5.3 ppb). Across the full year, random errors range between 5.3 and 7.4 ppb. Across all seasons, the emulator RMSE (5.3–7.4 ppb) remains substantially smaller than both the variability and the satellite retrieval noise. Even in seasons with the largest RMSE, the emulator error constitutes less than 50% of the background mole fraction standard deviation. We propose that this level of uncertainty makes the emulator appropriate for inverse modelling frameworks, because it is secondary to the primary sources of uncertainty, such as observational noise, and is relatively unbiased.

Figure 1 shows the spatial distribution of the background mole fraction contributions to the GOSAT observations, propagated from the boundaries using NAME and model predictions using GATES-Background. The figure also shows the corresponding

https://doi.org/10.5194/egusphere-2026-3361
Preprint. Discussion started: 28 September 2026

Table 1. Seasonal RMSE, bias, and random error for the predicted background mole fraction and the standard deviation for background methane mole fractions obtained from NAME and GATES-Background. The seasons are defined as January-February-March (JFM), April-May-June (AMJ), July-August-September (JAS), and October-November-December (OND).

Season	RMSE (ppb)	Bias (ppb)	σ_{error} (ppb)	SD_{NAME} (ppb)	$SD_{GATES-Background}$ (ppb)
JFM	7.4	-0.7	7.4	19.5	18.3
AMJ	6.6	-2.7	6.0	20.0	18.7
JAS	5.3	0.1	5.3	16.3	14.8
OND	6.4	-1.3	6.2	19.1	17.0
All	6.3	-1.1	6.2	18.9	17.6

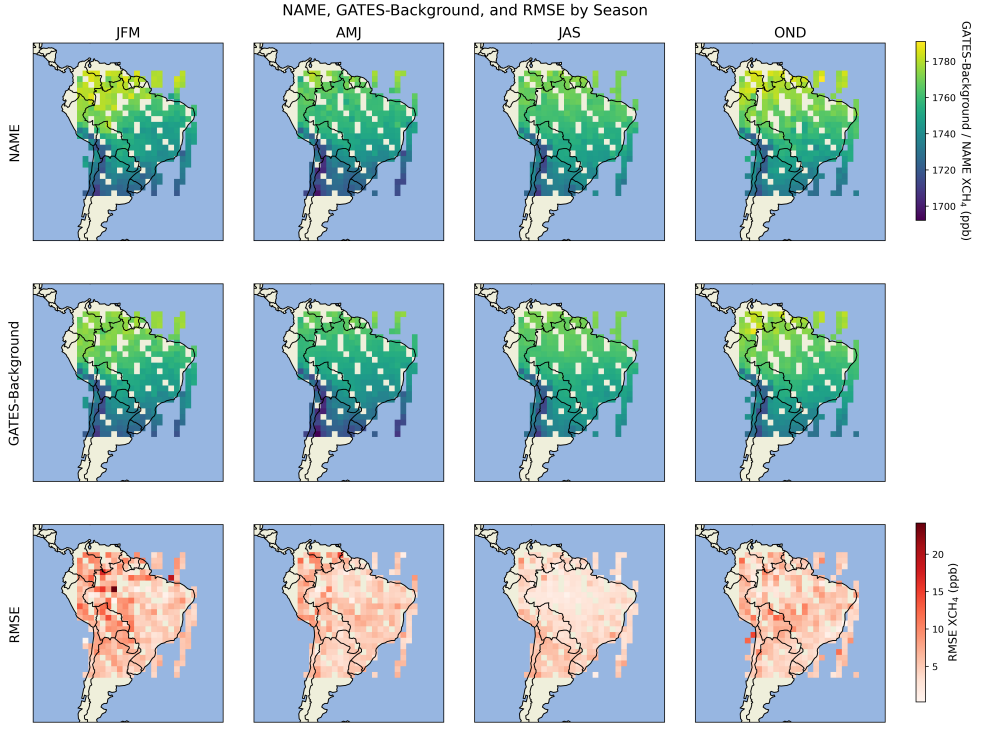

Figure 1. Seasonally-disaggregated spatial plots of methane mole fractions propagated from the domain boundary to the measurement point using NAME (top row), or predicted using GATES-Background (middle row). The bottom row shows RMSE values between the two across South America in 2017. Each column denotes a season, January-February-March (JFM), April-May-June (AMJ), July-August-September (JAS), and October-November-December (OND). For each pixel in the plot (with a resolution of $0.352^\circ \times 0.234^\circ, \approx 33 \times 25\,\text{km}$), all observations whose locations fell within the corresponding latitude and longitude bounds were assigned to that cell, and the mean value was computed.

https://doi.org/10.5194/egusphere-2026-3361
Preprint. Discussion started: 28 September 2026

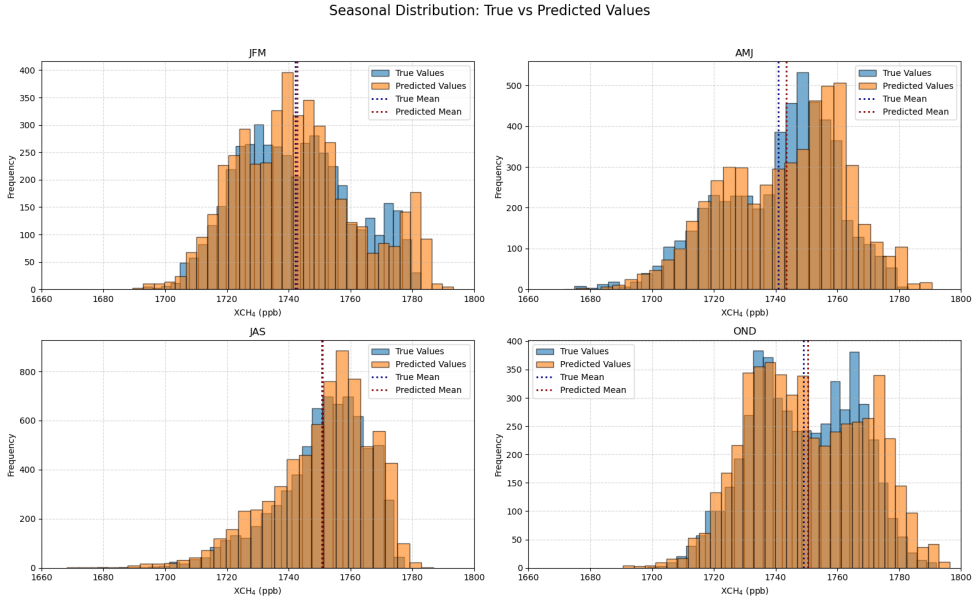

Figure 2. Histograms comparing the distribution of the NAME-predicted ("True") values and the predictions from the emulator in different seasons of 2017 for South America.

RMSE between the two. A comparison of the top two rows (true and prediction) reveals qualitatively similar spatial distributions across all seasons. The predicted fields closely resemble the NAME spatial patterns, indicating that the model successfully captures the large-scale seasonal and spatial variability, dominated by seasonal changes in boundary mole fractions and prevailing wind directions. Minor local discrepancies are present in some grid cells, which are highlighted more clearly in the corresponding RMSE maps (bottom row).

The RMSE maps show that the largest prediction errors tend to be spatially localized rather than widespread. Most of the domain is characterized by errors between approximately 3-5 ppb. Elevated errors occur primarily near the northern and north-western boundaries of the domain, particularly during January-February-March (approximately 15-20 ppb) and April-May-June (approximately 10-15 ppb). The larger error in these regions may partly reflect sparser data coverage, as shown in Fillola et al. (2026b), which can limit model training and constrain predictive accuracy. The western regions also show larger errors, which are likely due to modelling challenges associated with the complex topography in the Andes. These biases could be targeted in future studies by increasing the training dataset size over challenging regions, and/or algorithmic improvements.

Reduced spatial gradients correspond to a narrower distribution of background mole fractions, as illustrated by the seasonal histograms in Figure 2. The JAS season exhibits the most homogeneous spatial structure (low spatial gradients), resulting in the smallest spread of background values with true (NAME) and predicted (GATES-Background) standard deviations of 16.3 and 14.8 ppb, respectively. It also exhibits the lowest overall prediction error (RMSE of 5.3 ppb, compared to closer to 7 ppb on the

https://doi.org/10.5194/egusphere-2026-3361
Preprint. Discussion started: 28 September 2026

other seasons). Overall, the predicted histograms strongly resemble those calculated using NAME, capturing characteristics of the distribution such as the modes and the spread.

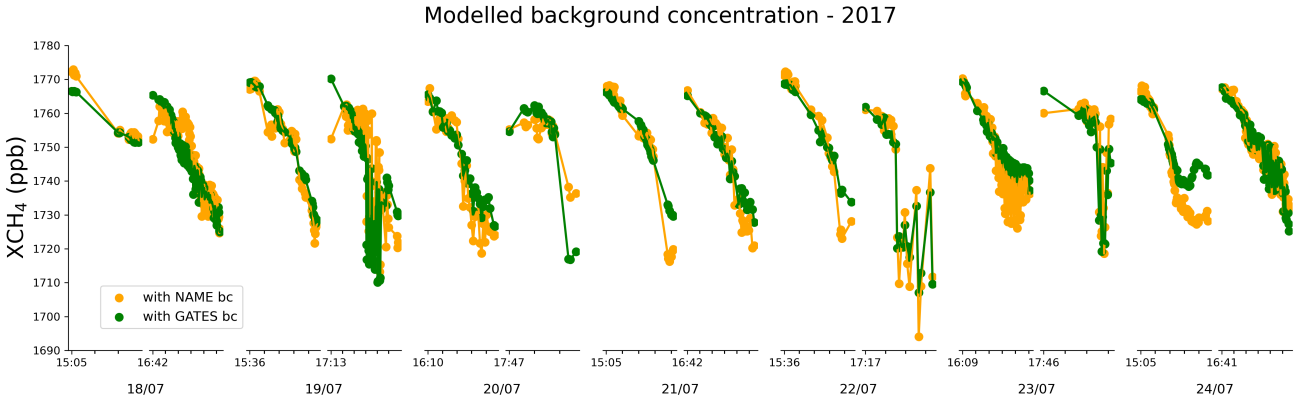

Figure 3. Modelled background mole fractions of CH_4 for different times in 2017 (randomly chosen) using the true mole fraction (NAME) compared to emulated predictions (GATES-Background).

Example GOSAT boundary condition XCH_4 timeseries for the emulator and NAME are shown in Figure 3, during the arbitrarily chosen period, 18th July to the 23rd July 2017. The predicted timeseries (GATES-Background) demonstrates overall agreement with NAME as the satellite traverses the continent during these days. However, some systematic differences in magnitude are evident; the NAME-based values exhibit greater short-timescale variability and regularly reach lower minimum mole fractions, while the GATES-Background-based values are generally smoother. This is typical behaviour when using MSE loss, which tends towards predictions close to the mean. Furthermore, GNNs are often associated with smoothed outputs (Rusch et al., 2023). The persistence of these differences across multiple days indicates that they arise from structural characteristics of the emulator rather than random variability.

Future work will focus on extending the framework in several directions. First, we aim to train and evaluate the model across additional geographic regions to assess its generalisability and robustness under a variety of different topographic and climatic regimes. Second, rather than predicting a single aggregated background value, we plan to model contributions from different sections of the domain boundary separately (e.g., northern, eastern, southern or western boundaries), enabling a more physically interpretable representation of background inflow, and more degrees of freedom that can be adjusted in an inversion (e.g. Lunt et al., 2016). Finally, we intend to integrate the background mole fraction emulator with existing GATES footprint architecture, creating a unified modelling framework capable of jointly predicting both local and background contributions to the total mole fraction.

https://doi.org/10.5194/egusphere-2026-3361
Preprint. Discussion started: 28 September 2026

4 Conclusions

In this study, we propose GATES-Background, a graph neural network (GNN) framework for emulating background mole fraction contributions to regional simulations of column mean mole fractions, for use in trace gas inversions of long-lived atmospheric gases. While previous approaches have primarily focused on predicting footprints to represent local flux contributions to the observed mole fractions, our work targets the background component, which constitutes a substantial portion of the total mole fraction and its variability. We demonstrate the effectiveness of the method for GOSAT methane observations over South America. The results show strong agreement with the numerical model-calculated values, capturing the major spatial and seasonal patterns with uncertainties smaller than the GOSAT XCH_4 uncertainty and the variability of the boundary condition mole fractions within the domain. This study represents an important step towards building a fully machine learning-based inverse modelling system.

Code and data availability. The code used to implement, train, and evaluate our model is present at DOI https://doi.org/10.5281/zenodo.21101081 (Fillola et al., 2026a). The full dataset of LPDM footprints (with the exception of the year 2017) generated with NAME can be found at DOI https://doi.org/10.5281/zenodo.16748754 (Fillola and Tunnicliffe, 2025). Additionally, LPDM footprints generated with NAME for 2017 can be found at DOI. https://doi.org/10.5281/zenodo.21031381 (Keshtmand and Tunnicliffe, 2026b). Also, the full dataset of the CAMS reanalysis products for the years 2014-2017 can be found at DOI https://doi.org/10.5281/zenodo.21035682 (Keshtmand and Tunnicliffe, 2026a). 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/UMG_Mk9 Met Office (2025a) and https://data.ceda.ac.uk/badc/name_nwp/data/global/UMG_Mk10 Met Office (2025b) under a UK Open Government Licence 3.0. The GOSAT XCH_4 retrievals are located at https://doi.org/10.5285/18ef8247f52a4cb6a14013f8235cc1eb (Parker and Boesch, 2020).

Author contributions. Conceptualization: EF, MR, RSR. Methodology: EF, MR, RSR. Investigation: NK. Software: NK, EF, JC. Visualization: NK, EF. Supervision: MR, RSR. Writing – original draft: NK. Writing – review and editing: NK, EF, JC, RSR, MR.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. NK and MR were funded by the Engineering and Physical Sciences Research Council (EPSRC) project, "Self-learning digital twins for sustainable land management" (EP/Y00597X/1). EF was funded by a Google Research PhD Scholarship. NK, EF, MR, RSR and JC received funding from Natural Environment Research Council (NERC) Pushing the Frontiers grant "Global Methane Flux Inference using Emulated Atmospheric Transport" (NE/Z504294/1).

https://doi.org/10.5194/egusphere-2026-3361
Preprint. Discussion started: 28 September 2026

References

Bacciu, D., Errica, F., Micheli, A., and Podda, M.: A gentle introduction to deep learning for graphs, Neural Networks, 129, 203–221, 2020.

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., et al.: Relational inductive biases, deep learning, and graph networks, arXiv preprint arXiv:1806.01261, 2018.

Fillola, E. and Tunnicliffe, R.: LPDM footprint dataset – Brazil, https://doi.org/10.5281/zenodo.16748754, [data set], 2025.

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, 2023.

Fillola, E., Clark, J., and Keshtmand, N.: GATES: A Graph-Neural-Network Atmospheric Transport Emulation System, https://doi.org/10.5281/zenodo.21101081, 2026a.

Fillola, E., Santos-Rodriguez, R., Tunnicliffe, R., Clark, J. N., Keshtmand, N., Ganesan, A., and Rigby, M.: Enabling fast greenhouse gas emissions inference from satellites with GATES: a Graph-Neural-Network Atmospheric Transport Emulation System, Geoscientific Model Development, 19, 1893–1915, 2026b.

Ganesan, A. L., Rigby, M., Lunt, M. F., Parker, R. J., Boesch, H., Goulding, N., Umezawa, T., Zahn, A., Chatterjee, A., Prinn, R. G., et al.: Atmospheric observations show accurate reporting and little growth in India's methane emissions, Nature Communications, 8, 836, 2017.

Gerrand, K., Fillola, E., Manning, A. J., Arduini, J., Krummel, P. B., Lunder, C. R., Mühle, J., O'Doherty, S., Park, S., Prinn, R. G., et al.: A Machine Learning Method for Estimating Atmospheric Trace Gas Concentration Baselines, EGUsphere, 2025, 1–20, 2025.

Hamazaki, T., Kaneko, Y., Kuze, A., and Kondo, K.: Fourier transform spectrometer for greenhouse gases observing satellite (GOSAT), in: Enabling sensor and platform technologies for spaceborne remote sensing, vol. 5659, pp. 73–80, SPIE, 2005.

He, T.-L., Dadheech, N., Thompson, T. M., and Turner, A. J.: FootNet: Development of a machine learning emulator of atmospheric transport, 2023.

Hu, L., Montzka, S. A., Lehman, S. J., Godwin, D. S., Miller, B. R., Andrews, A. E., Thoning, K., Miller, J. B., Sweeney, C., Siso, C., et al.: Considerable contribution of the Montreal Protocol to declining greenhouse gas emissions from the United States, Geophysical Research Letters, 44, 8075–8083, 2017.

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., et al.: The CAMS reanalysis of atmospheric composition, Atmospheric Chemistry and Physics, 19, 3515–3556, 2019.

Jones, A., Thomson, D., Hort, M., and Devenish, B.: The UK Met Office's next-generation atmospheric dispersion model, NAME III, in: Air pollution modeling and its application XVII, pp. 580–589, Springer, 2007.

Keisler, R.: Forecasting global weather with graph neural networks, arXiv preprint arXiv:2202.07575, 2022.

Keshtmand, N. and Tunnicliffe, R.: CAMS background methane mole fraction curtains – South America, https://doi.org/10.5281/zenodo.21035682, [data set], 2026a.

Keshtmand, N. and Tunnicliffe, R.: LPDM footprint dataset – Brazil 2017, https://doi.org/10.5281/zenodo.21031381, [data set], 2026b.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., et al.: GraphCast: Learning skillful medium-range global weather forecasting, arXiv preprint arXiv:2212.12794, 2022.

Lunt, M. F., Rigby, M., Ganesan, A. L., and Manning, A. J.: Estimation of trace gas fluxes with objectively determined basis functions using reversible-jump Markov chain Monte Carlo, Geoscientific Model Development, 9, 3213–3229, 2016.

Manning, A., O'Doherty, S., Jones, A., Simmonds, P., and Derwent, R.: Estimating UK methane and nitrous oxide emissions from 1990 to 2007 using an inversion modeling approach, Journal of Geophysical Research: Atmospheres, 116, 2011.

https://doi.org/10.5194/egusphere-2026-3361
Preprint. Discussion started: 28 September 2026

Met Office: Global NWP meteorological data for Met Office NAME dispersion model (Mk9: July 2015–2017) (Mk9), CEDA (Centre for Environmental Data Analysis) [data set], https://data.ceda.ac.uk/badc/name_nwp/data/global/UMG_Mk9, accessed: September 2025, 2025a.

Met Office: Global NWP meteorological data for Met Office NAME dispersion model (Mk10: June 2017–May 2022) (Mk10), CEDA (Centre for Environmental Data Analysis) [data set], https://data.ceda.ac.uk/badc/name_nwp/data/global/UMG_Mk10, accessed: September 2025, 2025b.

Parker, R. and Boesch, H.: University of leicester GOSAT proxy XCH4 v9. 0, (No Title), 2020.

Parker, R. J., Webb, A., Boesch, H., Somkuti, P., Barrio Guillo, R., Di Noia, A., Kalaitzi, N., Anand, J. S., Bergamaschi, P., Chevallier, F., et al.: A decade of GOSAT Proxy satellite CH 4 observations, Earth System Science Data, 12, 3383–3412, 2020.

Pfaff, T., Fortunato, M., Sanchez-Gonzalez, A., and Battaglia, P. W.: Learning mesh-based simulation with graph networks, arXiv preprint arXiv:2010.03409, 2020.

Rusch, T. K., Bronstein, M. M., and Mishra, S.: A survey on oversmoothing in graph neural networks, arXiv preprint arXiv:2303.10993, 2023.

Saunois, M., Martinez, A., Poulter, B., Zhang, Z., Raymond, P. A., Regnier, P., Canadell, J. G., Jackson, R. B., Patra, P. K., Bousquet, P., et al.: Global methane budget 2000–2020, Earth System Science Data, 17, 1873–1958, 2025.

Thompson, R. L. and Stohl, A.: FLEXINVERT: an atmospheric Bayesian inversion framework for determining surface fluxes of trace species using an optimized grid, Geoscientific Model Development, 7, 2223–2242, 2014.

Tunnicliffe, R. L., Ganesan, A. L., Parker, R. J., Boesch, H., Gedney, N., Poulter, B., Zhang, Z., Lavrič, J. V., Walter, D., Rigby, M., et al.: Quantifying sources of Brazil's CH 4 emissions between 2010 and 2018 from satellite data, Atmospheric Chemistry and Physics, 20, 13 041–13 067, 2020.

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.