https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


# GenGHG v1.0: A generative machine learning emulator of greenhouse-gas atmospheric transport around global emission hotspots

Zeyu Wang[1,2], Jieyi Wang[1,2], Hanyu Liu[3], Ziting Huang[3], Kewei Liang[4], Feng Zhang[1,2], and Scot M. Miller[3]

[1]School of Earth Sciences, Zhejiang University, Hangzhou 310058, China
[2]Zhejiang Provincial Key Laboratory of Geographic Information Science, Hangzhou 310058, China
[3]Department of Environmental Health and Engineering, Johns Hopkins University, Baltimore, MD, USA
[4]School of Mathematical Sciences, Zhejiang University, Hangzhou, 310027, China

**Correspondence:** Feng Zhang (zfcarnation@zju.edu.cn)

**Abstract.** Dense greenhouse-gas (GHG) observations, particularly from satellites, require fast atmospheric transport models that link surface emissions to these observations. We introduce GenGHG, a generative machine learning (ML) model that emulates footprints from the Stochastic Time-Inverted Lagrangian Transport (STILT) model. These footprints estimate how a unit of emissions would alter a downwind atmospheric measurement. Unlike existing deterministic ML emulators, which give a single prediction, GenGHG predicts an ensemble of plausible footprints under given meteorological forcing conditions to represent stochastic atmospheric transport. GenGHG retains ML-level computational efficiency, with each member generated in less than 2 s on a single GPU, and we evaluate its generative advantage over deterministic ML across a benchmark of 60 urban areas worldwide. Results show that GenGHG preserves footprint total mass substantially better than deterministic ML, with a mean relative bias of $+2.67\%$ compared with $-22.38\%$, and also better preserves the footprint-value distribution. Grid-level accuracy also shows that GenGHG better captures spatial footprint patterns, with larger gains in more dispersive transport regimes. We further test GenGHG's advantage in a synthetic inverse modeling experiment, where transport from GenGHG yields methane emission estimates that closely follow the STILT transport reference and outperform deterministic ML. Finally, we highlight GenGHG's compatibility with any global meteorology product at a 0.25 ° resolution. This flexibility, combined with its low computational cost, makes it straightforward to run footprints using multiple meteorology products and subsequently evaluate the possible effects of meteorological uncertainties.

## 1 Introduction

Effective climate mitigation requires reliable estimates of greenhouse-gas (GHG) emissions at decision-relevant scales (Janssens-Maenhout et al., 2020). Anthropogenic emissions are concentrated in cities, industrial regions, and energy-production basins, where localized enhancements can make a disproportionate contribution to national and global GHG budgets (Moran et al., 2018; Ramaswami et al., 2021; Lauvaux et al., 2022). Atmospheric observing networks have expanded from sparse surface networks to denser ground-based (Lauvaux et al., 2016; Karion et al., 2020), airborne (Duren et al., 2019) and satellite mea-

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


surements (Veefkind et al., 2012; Eldering et al., 2017), providing increasingly detailed constraints on these emission hotspots. Top-down inverse modeling based on atmospheric GHG concentrations can inform surface fluxes and help evaluate or refine bottom-up inventories (Jacob et al., 2022). This inverse modeling relies on atmospheric transport models to link measured signals to upwind emission sources. Eulerian models describe concentration evolution on fixed grids (e.g., Bey et al., 2001; Krol et al., 2005; Ganshin et al., 2012; Estrada et al., 2024), whereas Lagrangian models follow particles through meteorological fields to quantify source influence on observations (e.g., Stohl et al., 1998; Lin et al., 2003; Jones et al., 2007).

Among Lagrangian transport approaches, Lagrangian particle dispersion models (LPDMs) provide a practical framework for GHG source attribution by tracing observed particles back to potential upwind source regions (Stein et al., 2015). This backward calculation provides the footprint that links observed GHG signals to upwind surface emissions. For example, the Stochastic Time-Inverted Lagrangian Transport (STILT) model (Lin et al., 2003) has been applied from continental (e.g., Gerbig et al., 2003a; He et al., 2018) and regional (e.g., Reuter et al., 2014; Upton et al., 2026; Liu et al., 2026) to urban-scale (Kiel et al., 2021; Ohyama et al., 2023; Wilmot et al., 2024) GHG monitoring studies. Nonetheless, recent carbon dioxide ($CO_2$) and methane ($CH_4$) satellite observing systems are moving from sparse soundings toward denser sampling and higher-resolution imaging (Jacob et al., 2022; Tanimoto et al., 2025). As observation numbers continue to grow, the computational cost of full-physics particle simulations creates a scaling bottleneck for operational transport modeling and near-real-time inverse modeling of emission hotspots.

This scaling pressure has motivated machine learning (ML) approaches for faster and more operational links between atmospheric observations and emissions. ML has been used to exploit observed enhancements or plume structures for source attribution (e.g., Dumont Le Brazidec et al., 2023; Wang et al., 2026), and to emulate Lagrangian-based transport itself by learning the complex relationship between meteorological conditions and footprints (Fillola et al., 2023; He et al., 2025; Fillola et al., 2026). For transport emulation, ML models can reduce simulation time by orders of magnitude relative to LDPMs and have been developed for both ground-based, temporally dense observations and satellite, spatially dense observations. Despite these gains in computational efficiency, existing ML emulators still exhibit important limitations. Fillola et al. (2026) reported over-smoothed footprint predictions and used additional post-processing with manually chosen thresholds; while Dadheech and Turner (2026) reported difficulties in conserving total mass of the emulated footprint for column receptors. These limitations call for an ML framework that preserves physically meaningful footprint properties, while representing the stochastic nature of atmospheric transport.

Meteorological forcing is a primary control on atmospheric transport, in addition to the choice of transport model itself. Errors in meteorological fields can alter transport pathways and vertical mixing, thereby changing how surface emissions contribute to observed atmospheric signals (Lin and Gerbig, 2005; Gerbig et al., 2008). These forcing-dependent differences propagate through the transport model and affect both transport reliability and inverse modeling results (Deng et al., 2017; He et al., 2018; Munassar et al., 2023). For satellite observations, which sample specific times and locations, publicly available weather products are commonly used as meteorological forcing data. Differences among products can therefore express part of the transport uncertainty, as recognized in localized emission estimation studies (e.g., Nassar et al., 2017, 2021). Recent ML emulation work has begun to test model transferability across regions in the continental United States and across

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


two meteorological inputs (Dadheech and Turner, 2026), while systematic cross-product diagnosis of transport differences remains unexplored. This study therefore extends ML transport emulation to global emission hotspots and uses cross-product comparisons to diagnose meteorology-driven transport differences.

In this paper, we introduce GenGHG, a generative ML transport model for tracing atmospheric GHGs around anthropogenic emission hotspots worldwide. GenGHG retains the computational efficiency of ML models and advances GHG transport emulation in two aspects:

- First, GenGHG creates multiple stochastic realizations of each footprint, not just a single deterministic estimate as has been common in existing ML models. We find that this generative formulation yields substantially better emulation skill and stronger structural consistency with the full-physics reference.
- Second, GenGHG is built on global benchmark and uses medium-resolution weather products as meteorological forcing. This design enables flexible application to satellite observations over different emission hotspots, and also the low-cost diagnosis of transport variability across different weather products.

These two aspects of GenGHG are evaluated through the following analyses. We first test the generative formulation on a global benchmark by assessing structural consistency in total footprint and per-grid-cell prediction skill (Section 3.1). We then examine the operational applicability of GenGHG through out-of-benchmark applications, including GHG emission inverse modeling in an independent case study (Section 3.2) and transport-pattern diagnosis driven by untrained weather products (Section 3.3).

## 2 Methods and materials

### 2.1 From Lagrangian footprints to generative emulation

Atmospheric observations respond to surface fluxes from many upwind locations and times. In a linear transport formulation, surface fluxes over the $n$ source grid cells are represented by the source vector $\boldsymbol{s}$ (dimensions $n \times 1$), and the $m$ atmospheric observations are represented by the receptor vector $\boldsymbol{y}$ ($m \times 1$). The transport operator $\mathbf{H}$ ($m \times n$), which represents the source–receptor relationships, links the two through $\boldsymbol{y} = \mathbf{H}\boldsymbol{s} + \boldsymbol{\epsilon}$ (Gerbig et al., 2003b), where $\boldsymbol{\epsilon}$ ($m \times 1$) is the residual error. Each element of $\mathbf{H}$ describes how emissions from one surface element contribute to one observation. In LDPMs, each row of $\mathbf{H}$ corresponds to a single atmospheric observation and can be explicitly computed by tracking particles backward from a receptor through meteorological fields. This term is called a footprint, denoted $\boldsymbol{h}$ ($n \times 1$), and is the basic emulation target of GenGHG in this study.

GenGHG retains the Lagrangian transport view, but learns the footprint as a whole sample through its generative evolution, rather than explicitly simulating particle trajectories in a physical system. The common idea of LDPMs and GenGHG can be written as a general transport-state evolution under meteorological conditions $\boldsymbol{c}$:

$$\frac{\mathrm{d}\boldsymbol{q}_t}{\mathrm{d}t} = \boldsymbol{u}(\boldsymbol{q}_t, t; \boldsymbol{c}), \tag{1}$$

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


where $\boldsymbol{q}_t$ is the evolving transport state at time $t$, and $\boldsymbol{u}$ denotes the velocity field that moves this state. LPDMs and GenGHG define these terms differently:

- In LPDMs, $\boldsymbol{q}_t$ represents the positions of particles released from a receptor and evolved backward in physical time $t$. The velocity field $\boldsymbol{u}$ represents resolved atmospheric motion, which is combined with parameterized stochastic turbulence to simulate particle dispersion. The footprint $\boldsymbol{h}$ is obtained by aggregating particle residence time within upwind source grid cells.

- In GenGHG, $\boldsymbol{q}_t$ is the evolving footprint state $\boldsymbol{h}_t$, where $t$ is the normalized generative time between the initial noise and the target footprint. A neural network $\theta$ learns a latent-space flow, described by the velocity term $\boldsymbol{u}_\theta$, that moves $\boldsymbol{h}_t$ along this generative path.

### 2.2 GenGHG formulation and architecture

We use flow matching (FM) (Lipman et al., 2022) as the generative formulation for GenGHG. FM learns a flow field that moves samples from a prior distribution toward a target data distribution, which is naturally aligned with the footprint evolution equation described in Eq. 1. Recent work has also applied FM to numerical weather prediction in atmospheric science (Couairon et al., 2026) and human mobility flow prediction in geographical science (Xu et al., 2026), supporting its use for scientific prediction tasks. Based on this FM formulation, GenGHG learns the distribution of footprints given meteorological conditions, denoted $P(\boldsymbol{h} \mid \boldsymbol{c})$. The full GenGHG workflow is summarized in Fig. 1 and described below.

**Footprint ($\boldsymbol{h}$):** GenGHG generates footprints over a two-dimensional geographic domain (the right panel of Fig. 1). Their spatial pattern and magnitude describe where, and to what extent, surface fluxes influence a receptor observation. Specifically, each grid-level element of $\boldsymbol{h}$ gives the concentration response at the receptor to a unit surface flux from one source grid cell, with units of ppm $(\mu\mathrm{mol\ m^{-2}\ s^{-1}})^{-1}$. Because footprint values are highly skewed in original unit, model training and skill evaluation are performed in logarithmic space.

**Meteorological forcing ($\boldsymbol{c}$):** Meteorological forcing fields are used as the primary conditions that drive footprint generation, supplemented by receptor-relative information (left panel of Fig. 1; Table 1). These fields include multi-pressure-level winds, derived variables, surface pressure and planetary boundary-layer height, sampled at the observation time and every 6 hours over the 24-hour backward window. Static receptor-relative fields encode the distance from each source grid cell to the receptor.

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


**Table 1. Meteorological forcing and receptor-relative variables used by GenGHG.** Multi-temporal denotes fields sampled at the observation time and every 6 hours over the 24-hour backward window. Multi-level denotes the near-surface level (10 m) and pressure levels at 925, 850, 700, 600 and 500 hPa.

| Conditioning field | Variables | Temporal and vertical support |
|---|---|---|
| Wind field | Zonal wind $u$, meridional wind $v$ | Multi-temporal; multi-level |
| Derived transport variables | Wind speed, ideal backward Gaussian plume (He et al., 2025) | Multi-temporal; multi-level |
| Surface and boundary-layer state | Surface pressure, planetary boundary-layer height | Multi-temporal; single-level |
| Receptor-relative position | Euclidean distance and normalized distance to receptor | static |

**Generative training and ensemble prediction:** During training, GenGHG learns the velocity field that connects a simple prior distribution to the target footprint distribution (middle panel of Fig. 1). Each training path pairs a Gaussian prior sample $\boldsymbol{h}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, which contains no prescribed transport structure, with a target footprint $\boldsymbol{h}_1$. For a normalized generative time $t \sim \mathcal{U}(0,1)$, the interpolated state is $\boldsymbol{h}_t = (1-t)\boldsymbol{h}_0 + t\boldsymbol{h}_1$. Along this path, the target velocity is $\boldsymbol{h}_1 - \boldsymbol{h}_0$, and the neural network $\theta$ is trained to predict this velocity $\boldsymbol{u}$ from the interpolated state ($\boldsymbol{h}_t$, $t$) and the meteorological condition $\boldsymbol{c}$ using the conditional flow-matching loss

$$\mathcal{L}_{\mathrm{Gen}} = \mathbb{E}_{\boldsymbol{h}_1, \boldsymbol{h}_0, t, \boldsymbol{c}} \left[ \| \boldsymbol{u}_\theta(\boldsymbol{h}_t, t; \boldsymbol{c}) - (\boldsymbol{h}_1 - \boldsymbol{h}_0) \|_2^2 \right]. \quad (2)$$

At the training stage, each footprint simulated by a physical LDPM (detailed in Section 2.3) is referred to as a sample $\boldsymbol{h}_1$ from the target distribution, defining one path from the prior to the data. Intermediate times $t$ are sampled with a cosine schedule, following the practice of allocating more steps to low-noise states to improve generation quality (Nichol and Dhariwal, 2021).

At the prediction stage, the trained network generates a footprint for a given receptor and meteorological condition $\boldsymbol{c}$ by starting from a Gaussian prior sample $\boldsymbol{h}_0$. The learned velocity field $\boldsymbol{u}_\theta(\boldsymbol{h}_t, t; \boldsymbol{c})$ is integrated from $t = 0$ to $t = 1$ using the same cosine schedule with 50 discrete steps. Repeating this procedure with different prior samples yields an ensemble of predicted members from the learned ML distribution in log-transformed space. The final footprint distribution is represented by transforming these members back to the original units.

The ensemble mean in the original units is used as the final GenGHG footprint for total mass analysis, structural evaluation and inverse modeling applications. For ML performance assessment, however, the log-space ensemble mean more directly represents prediction skill in the learning space, and is therefore used for checkpoint selection, grid-level accuracy evaluation and visualization.

**Neural network:** The neural network consists of a primary transport module ($\theta$) (Fig. 1 middle-bottom part) and a subgrid adapter. The primary transport module first encode the meteorological conditions $\boldsymbol{c}$, the intermediate time $t$ and the interpolated state $\boldsymbol{h}_t$ into a shared latent space; it then learns latent footprint generation by optimizing $\boldsymbol{u}_\theta$ with the flow-matching loss defined above. We implement this module with a transformer architecture (Vaswani et al., 2017), whose self-attention mechanism represents interactions between the receptor and source-grid cells across the transport domain. The subgrid adapter maps the generated latent footprint from the $0.25°$ condition grid back to the native $1/32°$ source grid, providing the spatial

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


support needed to resolve anthropogenic emission hotspots. This adapter is implemented as a variational autoencoder (VAE) (Kingma and Welling, 2013), trained separately and then fixed during primary-module training, following common practice in latent diffusion and FM models (Rombach et al., 2022; Dao et al., 2023).

Detailed model configuration and training settings are provided in Appendix A. The same architecture, without the embedding for $\boldsymbol{h}_t$ and $t$, can also be trained deterministically to regress meteorological conditions to a single footprint. We use this deterministic version as the ML baseline, denoted DeterGHG.

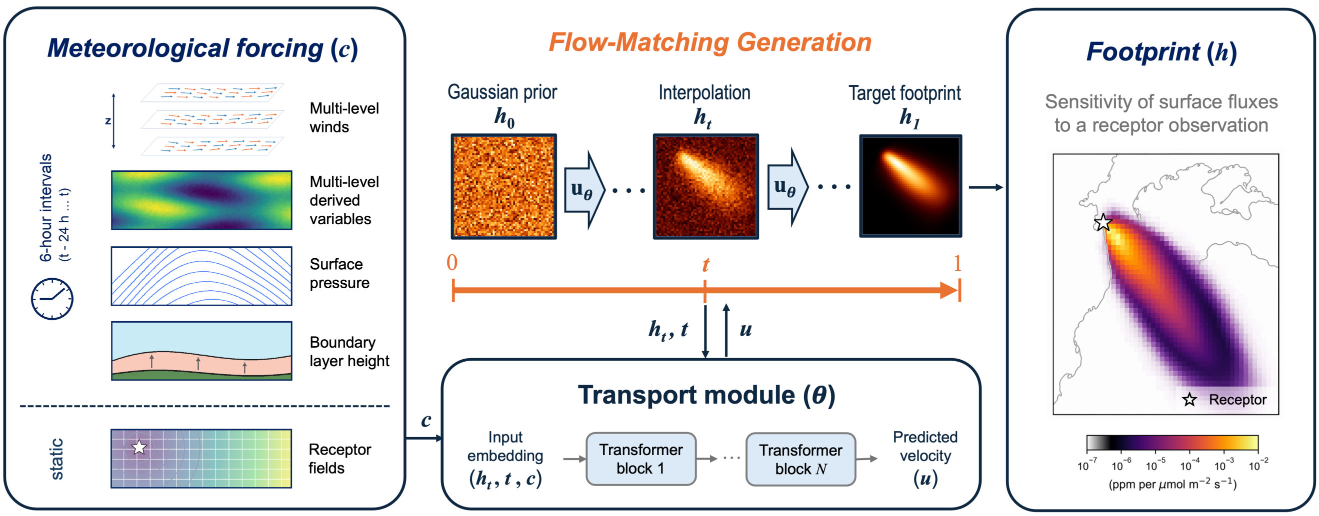


**Figure 1. GenGHG framework for generative footprint emulation.** Conditioned on meteorological forcing fields, the transport module moves a Gaussian prior toward a target footprint through flow-matching generation.

### 2.3 STILT reference footprint simulations

We use STILT (Lin et al., 2003) to generate the full-physics footprints used as reference targets for GenGHG training and evaluation. Several LPDMs can be used for atmospheric dispersion and footprint estimation, including the FLEXible PARTicle dispersion model (FLEXPART) and the Numerical Atmospheric-dispersion Modelling Environment (NAME). We chose STILT because its time-inverted formulation is well suited to GHG source attribution from atmospheric observations. It has also been broadly used in urban-scale emission studies, consistent with the spatial scale targeted by GenGHG.

#### 2.3.1 Base configuration

For this study, we configured STILT to simulate near-field GHG transport around local emission hotspots observed by satellite column receptors. The key settings and their rationale are as follows:

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


1. **Near-field footprint domain.** Each footprint represents 24 hours of backward transport over a 6°-latitude × 6°-longitude spatial domain and is aggregated onto a $1/32°$ (∼3-km) source grid. This setup targets near-field source contributions while resolving fine-scale emission structures.

2. **Column receptor and particle number.** Each column receptor is represented by 8,000 particles released across prescribed pressure levels and tracked backward in time. This relatively large ensemble provides sufficient sampling across pressure levels, reduces particle sampling noise and supports stable footprint estimates for the physical reference dataset.

3. **Pressure weighting.** Particle residence times near the surface are converted into surface influence and vertically integrated across pressure levels using the X-STILT configuration for satellite observations (Wu et al., 2018). Sensor-specific corrections are not applied because the particles contributing to near-field surface influence remain mainly in the lower troposphere, where GHG satellite averaging kernels are close to unity.

#### 2.3.2 Mass-conservation screening

Mass conservation means that particles are neither artificially created nor lost during transport. When this condition is violated, the near-surface influence of particles can be artificially diluted or concentrated, leading to biased footprints. Such biases are problematic for inverse modeling because inferred fluxes may compensate for transport artifacts rather than emissions. In STILT, such violations can arise when instantaneous, coarse-resolution meteorological fields are interpolated onto the grid used for particle transport, especially when the flow field is not mass-coupled to the transport calculation (Lin et al., 2003). Nehrkorn et al. (2010) addressed this problem using dynamically downscaled, mass-coupled meteorology in Weather Research and Forecasting (WRF)–STILT. However, repeating such a coupled setup at global scale for many receptor times and weather products is not practical for building a globally applicable GenGHG training benchmark.

We drove STILT with globally available, medium-resolution meteorological products and therefore applied additional quality control to each simulated footprint. As detailed in Appendix B, the STILT mass-conservation diagnostic "destruction of mass" (dmas) was used to classify each footprint simulation as pass, fail or badly fail. Only simulations classified as pass were retained in the benchmark for GenGHG training and evaluation. Simulations classified as badly fail were used as a stress-test set to evaluate whether GenGHG retains footprint structure learned from the benchmark in cases where STILT exhibits mass violations.

### 2.4 Benchmark and weather products

We constructed a benchmark of 47,998 footprints for receptors located around 60 urban emission hotspots worldwide from August 2019 to February 2025 (Fig. 2). The selected hotspots span all inhabited continents and a broad range of terrain settings (see Appendix C). This spatial coverage is paired with multi-year temporal sampling: receptor dates were sampled across months and split into training, validation and test sets at an 8:1:1 ratio. Receptor hours within each date were set to afternoon local time (approximately 13:00) and 00:00 UTC to cover satellite-relevant sampling windows across global time zones. For

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


each emission hotspot, receptors were placed at the city center and at eight surrounding land locations offset by $\pm 1.5°$, allowing transport around each hotspot to be sampled from multiple nearby directions.

Benchmark STILT simulations were driven by Global Forecast System (GFS) data from the National Oceanic and Atmospheric Administration (NOAA) (NOAA NCEP, 2015b), using lead time 0 fields at a native spatial resolution of 0.25°. GenGHG was also applied to the benchmark test set using three additional publicly available 0.25° weather products: Global Final Analysis data (FNL) (NOAA NCEP, 2015a), also from NOAA; Integrated Forecasting System data (IFS) (European Centre for Medium-Range Weather Forecasts, 2026) from the European Centre for Medium-Range Weather Forecasts (ECMWF); and the fifth-generation ECMWF atmospheric reanalysis (ERA5) (Copernicus Climate Change Service, Climate Data Store, 2023b, a), also from ECMWF. Together with the benchmark GFS forcing, these products support matched footprint prediction samples for diagnosing differences caused by meteorological forcing.

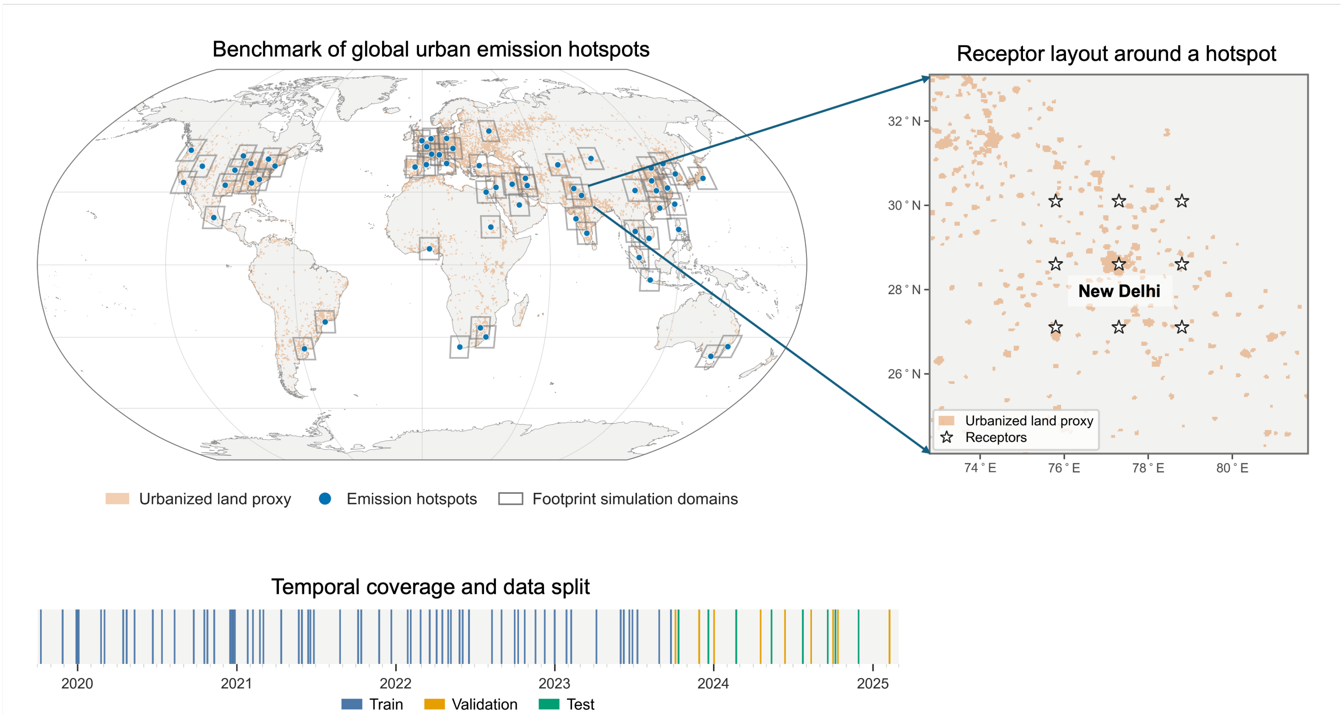


**Figure 2. Global urban emission-hotspot benchmark.** Receptors cover 60 urban emission hotspots worldwide and multi-year time coverage; local receptor placement is illustrated for New Delhi.

### 2.5 Synthetic inverse modeling experiment design

Observing system simulation experiments (OSSEs) use synthetic observations with a known truth to evaluate observing configurations and inverse modeling methods (Turner et al., 2018; Zhang et al., 2025). We used this framework to test whether the emulated footprints can serve as the transport term in inverse modeling under controlled conditions. The experiment was

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


conducted over the main Permian Basin domain (30–34°N, 105–100°W), a major oil-and-gas methane source region. This region is located outside the benchmark urban receptors, contains spatially complex local emissions, and has been widely used as a case study for top-down inverse modeling methods (e.g., Cusworth et al., 2021; Varon et al., 2023).

Synthetic observations were generated as daily $XCH_4$ snapshots from 24 May to 4 June 2025. For each day, receptor availability was based on the actual spatio-temporal sampling of blended satellite observations from the Tropospheric Monitoring Instrument (TROPOMI) and Greenhouse Gases Observing Satellite (GOSAT) (Balasus et al., 2023), assigned to 13:30 local time (18:00 UTC) and aggregated to $1/8°$ spatial resolution. Given this spatio-temporal observation availability, synthetic observations ($\boldsymbol{y}^{\mathrm{syn}}$) were generated by applying the reference transport operator ($\mathbf{H}_{\mathrm{true}}$) to a prescribed true flux field ($\boldsymbol{s}$), with additive observation noise ($\boldsymbol{\epsilon}$):

$$\boldsymbol{y}^{\mathrm{syn}} = \mathbf{H}_{\mathrm{true}}\boldsymbol{s}_{\mathrm{true}} + \boldsymbol{\epsilon}. \tag{3}$$

We used STILT footprints driven by GFS meteorology and passing the mass-conservation screening as $\mathbf{H}_{\mathrm{true}}$. Days with more than 100 valid observations were retained, yielding 9 daily inverse model runs with varying transport patterns and observation sampling. Synthetic $XCH_4$ enhancements were attributed only to oil-and-gas emissions within the domain and within the 24 h transport window, with homogeneous background and boundary contributions. We considered one noise-free case and three noisy cases with independent Gaussian observation noise, $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2_{\mathrm{obs}})$, with $\sigma_{\mathrm{obs}}$ =5, 8 and 10 ppb.

The inverse modeling settings were designed to isolate the role of the transport operator $\mathbf{H}$, which was supplied by STILT, GenGHG or DeterGHG footprints. The true surface flux field $\boldsymbol{s}$ was taken from a measurement-based oil-and-gas methane inventory (Omara et al., 2024), whereas the prior surface flux field $\boldsymbol{s}_{\mathrm{p}}$ was taken from the US Environmental Protection Agency (EPA) anthropogenic greenhouse-gas emission inventory (Maasakkers et al., 2016). The prior and true flux fields differ substantially in both total magnitude and hotspot structure, providing a biased prior for the inverse model to correct. Posterior fluxes were optimized separately for each daily satellite-overpass snapshot using the Bayesian inverse model described in Appendix D.

## 3 Results

### 3.1 Footprint emulation performance

GenGHG retains the computational efficiency expected from ML transport emulation. On an NVIDIA RTX 3090 GPU, one ensemble member generated with 50 integration steps takes less than 2 s, and both ensemble members and multiple receptors can be processed in parallel. Unless otherwise stated, the following analyses use a 30-member GenGHG ensemble. To isolate the contribution of probabilistic generation, we compare GenGHG with its deterministic counterpart, DeterGHG, which uses the same neural architecture and represents the baseline ML emulation capability without ensemble generation.

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


#### 3.1.1 Total mass and structural consistency

On the constructed global benchmark, we first evaluated whether the emulated footprints preserve the total structural consistency with the physical references given by STILT footprints. The analysis focuses on total mass, defined for each footprint as the sensitivity summed over all surface grid cells, together with the overall distribution of grid-level values and receptor-centered spatial scaling.

As shown in Fig. 3, GenGHG closely conserves total mass, with a mean relative bias of $+2.67\%$; the deterministic baseline instead shows a systematic negative bias of $-22.38\%$. The grid-level probability density function further shows that GenGHG reproduces the value range of the physical simulations across several orders of magnitude, including the high-sensitivity values below $10^{-1}$ that shape near-receptor influence, without generating the anomalous values above $10^{-1}$ produced by DeterGHG. The receptor-centered radial profile further evaluates spatial consistency by summarizing mean footprint sensitivity as a function of distance from the receptor. Physical footprints exhibit an approximate power-law decay in mean sensitivity, $\overline{H}(r) \propto r^{-\beta}$, where $r$ is the distance from the receptor and $\beta > 0$ is the decay exponent, reflecting distance-dependent sensitivity dilution. Both GenGHG and DeterGHG broadly reproduce this macroscopic scaling. However, DeterGHG underestimates sensitivity over the 10–300 km range, which accounts for more than $99\%$ of the surface area, explaining its negative mass bias. Within 10 km of the receptor, both GenGHG and DeterGHG show sensitivity excess, although the excess is more modest for GenGHG. The closer agreement in footprint-value distribution and radial scaling explains the low mass bias of GenGHG.

These results show that GenGHG better preserves the structural consistency to physical reference by learning a conditional distribution rather than a single estimate. Two factors explain this difference. First, DeterGHG provides only a point estimate in logarithmic space, which is more prone to biased structural collapse after transformation back to footprint original units. Second, probabilistic generation reduces over-smoothing and retains transport variability, consistent with recent Earth-system ML studies (Alet et al., 2025; Lang et al., 2026).

The preceding analyses use quality-controlled STILT simulations as physical references. We then assess whether GenGHG remains robust beyond the reference regime by applying the same structural metrics to STILT runs flagged with mass-violation problems, as defined by the dmas screening in Section 2.3.2. In this stress-test set ($n = 1{,}056$), the physical footprints exhibit an elevated frequency of anomalous values ($> 10^{-1}$) and radial profiles that lie markedly above the characteristic scaling of the benchmark simulations (Fig. 3). By contrast, the ML emulators preserve the footprint-value distribution and radial scaling learned from the benchmark, diverging from the flagged STILT simulations. GenGHG still retains its advantage over DeterGHG, preserving the structural benefits identified in the quality-controlled benchmark test set. These results indicate that statistically trained ML emulators can transfer the total-footprint structure learned from the benchmark to challenging application cases, including cases where STILT simulations are affected by mass violations. This robustness supports the operational use of GenGHG for rapid footprint generation from medium-resolution weather products.

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


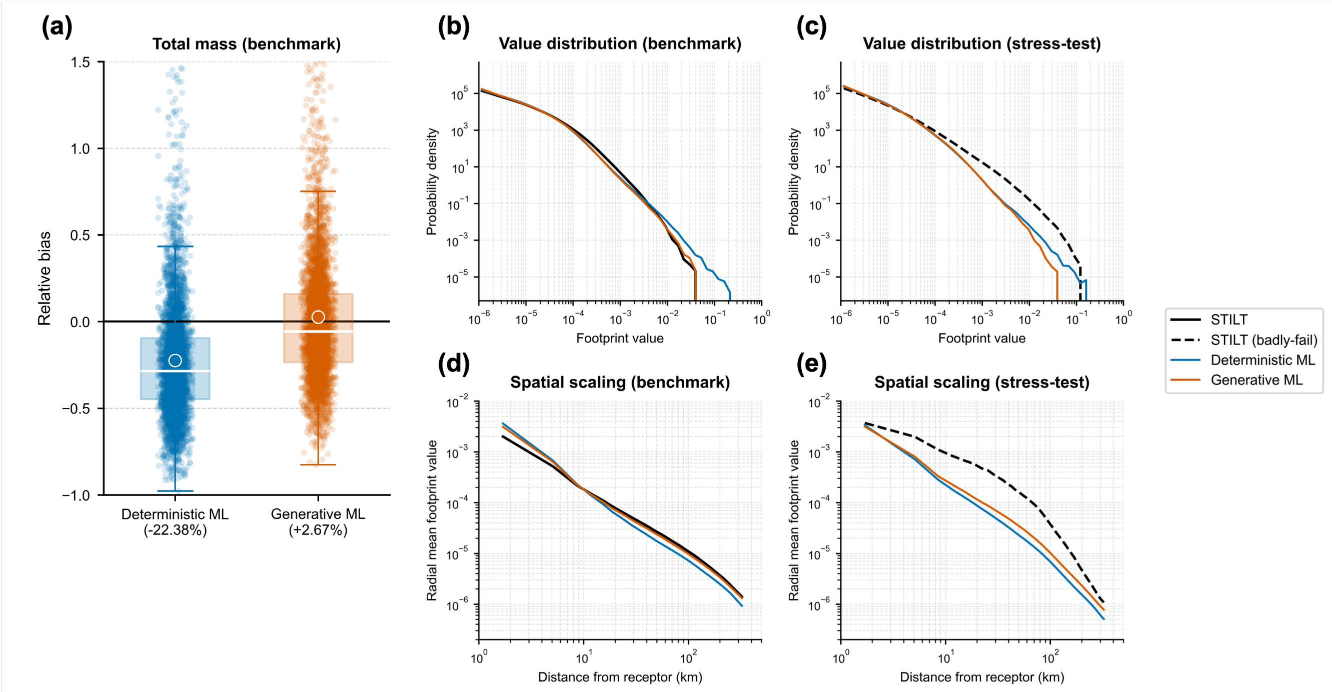


**Figure 3. Footprint mass and structural consistency.** (a) Relative bias in total mass, defined as footprint sensitivity summed over all surface grid cells. (b,c) Footprint-value distributions for the benchmark and stress-test cases. (d,e) Receptor-centered radial scaling for the benchmark and stress-test cases.

### 3.1.2 Grid-level accuracy

We further quantified GenGHG skill at the per-grid-cell level in logarithmic units. Across the full test set (Table 2), GenGHG reduces the mean squared error (MSE) from 0.785 to 0.683, lowers the root-mean-square error (RMSE) from 0.831 to 0.776, and increases the Pearson correlation coefficient (PCC) from 0.840 to 0.860 relative to DeterGHG, while giving a slightly higher mean absolute error (MAE; 0.290 versus 0.282). For the highly dispersive subset, defined as footprints with non-zero sensitivities covering more than 30% of the near-field domain (391 samples; $\sim 8\%$ of the test set), both ML emulators show lower accuracy, indicating the greater difficulty of emulating complex transport footprints. In this harder subset, however, the advantage of GenGHG over DeterGHG is clearer: GenGHG reduces RMSE from 1.123 to 1.040 and MSE from 1.364 to 1.167, and the MAE advantage also shifts to GenGHG.

The case examples in Fig. 4 complement the grid-level metrics by showing how these accuracy differences appear in spatial footprint patterns. In case 1, the receptor is located over Buenos Aires and represents a relatively compact coastal transport case. In case 2, the receptor is located between Dallas and Houston; this footprint is drawn from the highly dispersive subset and spans a broader influence area shaped by stronger changes in the wind field. In both cases, DeterGHG captures the dominant backward-transport direction, but compresses the broader STILT reference footprint into a smoother and narrower field.

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


By contrast, individual GenGHG members represent a diverse yet physically plausible set of transport pathways, and their ensemble mean more closely follows the dispersive pattern of the STILT footprint. Together with the quantitative evaluation, these examples show that GenGHG better preserves both the dominant footprint signal and variability, particularly under more dispersive and temporally changing wind conditions.

Beyond footprint dispersiveness itself, we examined whether GenGHG performance depends on the geographical setting of the receptors. Terrain complexity is a recognized factor affecting LPDM transport simulations (Brioude et al., 2012), and we therefore grouped the benchmark emission-hotspot locations into five topographic classes (Appendix C). The fraction of STILT simulations passing the mass-conservation screening decreases markedly from lowland and inland plain settings to mountainous regions (Fig. 5). Within the quality-controlled test set, however, GenGHG performance varies more modestly across terrain classes, with mean RMSE ranging from 0.735 over inland plains to 0.870 over mountainous regions. This terrain-related difference is smaller than the accuracy loss associated with highly dispersive transport patterns, indicating that transport complexity itself is the stronger challenge for emulation. The terrain sensitivity check therefore provides evidence that GenGHG remains broadly stable across the geographical settings represented in the global benchmark.

**Table 2. Comparison of deterministic and generative ML footprint prediction skill.** Metrics are averaged in the model learning space, i.e., the logarithm of the physical footprint unit. The dispersive subset contains footprints with non-zero sensitivities covering more than 30% of the near-field domain.

| Model | MSE ↓ | MAE ↓ | RMSE ↓ | PCC ↑ |
|---|---|---|---|---|
| *All test set* ($n = 4{,}630$) | | | | |
| DeterGHG | 0.785 | **0.282** | 0.831 | 0.840 |
| GenGHG | **0.683** | 0.290 | **0.776** | **0.860** |
| *Dispersive subset* ($n = 391$) | | | | |
| DeterGHG | 1.364 | 0.582 | 1.123 | 0.841 |
| GenGHG | **1.167** | **0.572** | **1.040** | **0.860** |

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


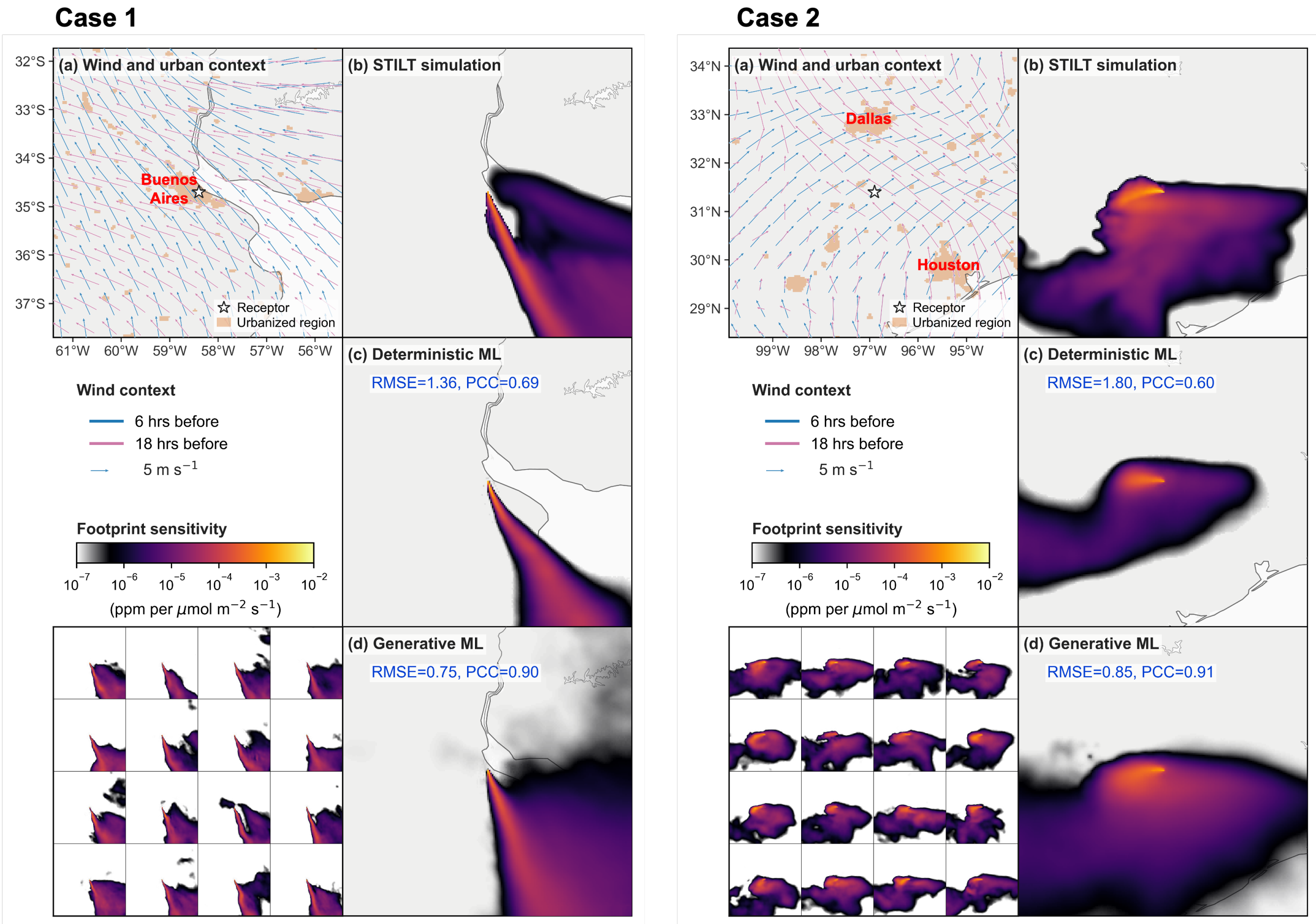


**Figure 4. Example transport footprints.** Two receptor cases are shown. For each case, panels show (a) wind and urban context, (b) the STILT simulation used as the reference, (c) the deterministic ML prediction and (d) the generative ML prediction. In panel (d), the main map shows the GenGHG ensemble mean, and the 16 insets on the left show randomly selected ensemble members.

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


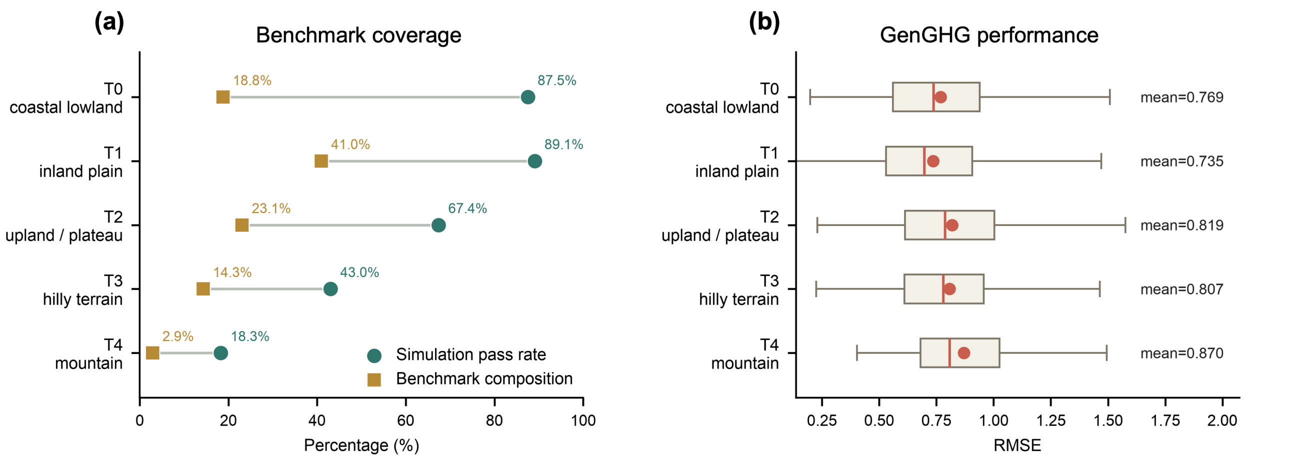


**Figure 5. Terrain-dependent benchmark coverage and performance.** (a) STILT simulation pass rate and benchmark composition across five topographic classes. (b) GenGHG emulation performance distributions across the same terrain classes.

### 3.2 Emission inverse modeling with emulated footprints

We assessed the use of GenGHG footprints as the forward transport operator in the synthetic inverse modeling experiment described in Section 2.5. The effectiveness of the emulated footprints in an inverse modeling setting was evaluated by testing whether the posterior recovered the prescribed true emissions and reduced the model–observation mismatch. In the idealized setup, with noise-free observations and STILT footprints used as the "true" transport operator, the posterior recovers the main emission structure well (Fig. D1). This result confirms that the synthetic observing network provides meaningful spatial and temporal constraints, and provides a baseline for comparing posterior differences caused by the transport operator.

The inverse modeling statistics in Table D1 show that GenGHG produces posterior emissions that more closely follow the STILT reference than DeterGHG, and this hierarchy is retained across observation-noise levels. Fig. 6 shows the 5 ppb noise case as an example. The daily posterior-to-prior ratio of total emission is $2.64 \pm 0.95$ for GenGHG, compared with $2.37 \pm 0.62$ for the STILT reference, whereas DeterGHG gives a larger and more variable correction of $3.45 \pm 2.28$. The day-to-day results also show extreme corrections when DeterGHG is used as the transport operator, whereas GenGHG remains within a more reasonable range on most daily inverse model runs. Spatially, GenGHG remains close to the STILT posterior in both emission pattern and hotspot intensity, while DeterGHG produces stronger spatial artefacts and lower agreement.

Beyond the posterior emission fields, the model–observation mismatch diagnostics further support the effectiveness of the GenGHG transport operator. In the 5 ppb noise case, the GenGHG posterior shifts the $XCH_4$ residuals from strongly negative values towards near zero, from a prior value of $-2.67_{\pm 6.44}$ ppb to $-0.37_{\pm 5.09}$ ppb, close to the STILT reference ($-0.14_{\pm 4.80}$ ppb). By contrast, the DeterGHG posterior retains a larger negative mismatch of $-0.70_{\pm 5.16}$ ppb, indicating that deterministic transport emulation leaves a stronger residual bias in observation space. These results indicate that footprints from GenGHG can support inverse modeling studies with substantially less posterior bias than deterministic transport emulation.

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


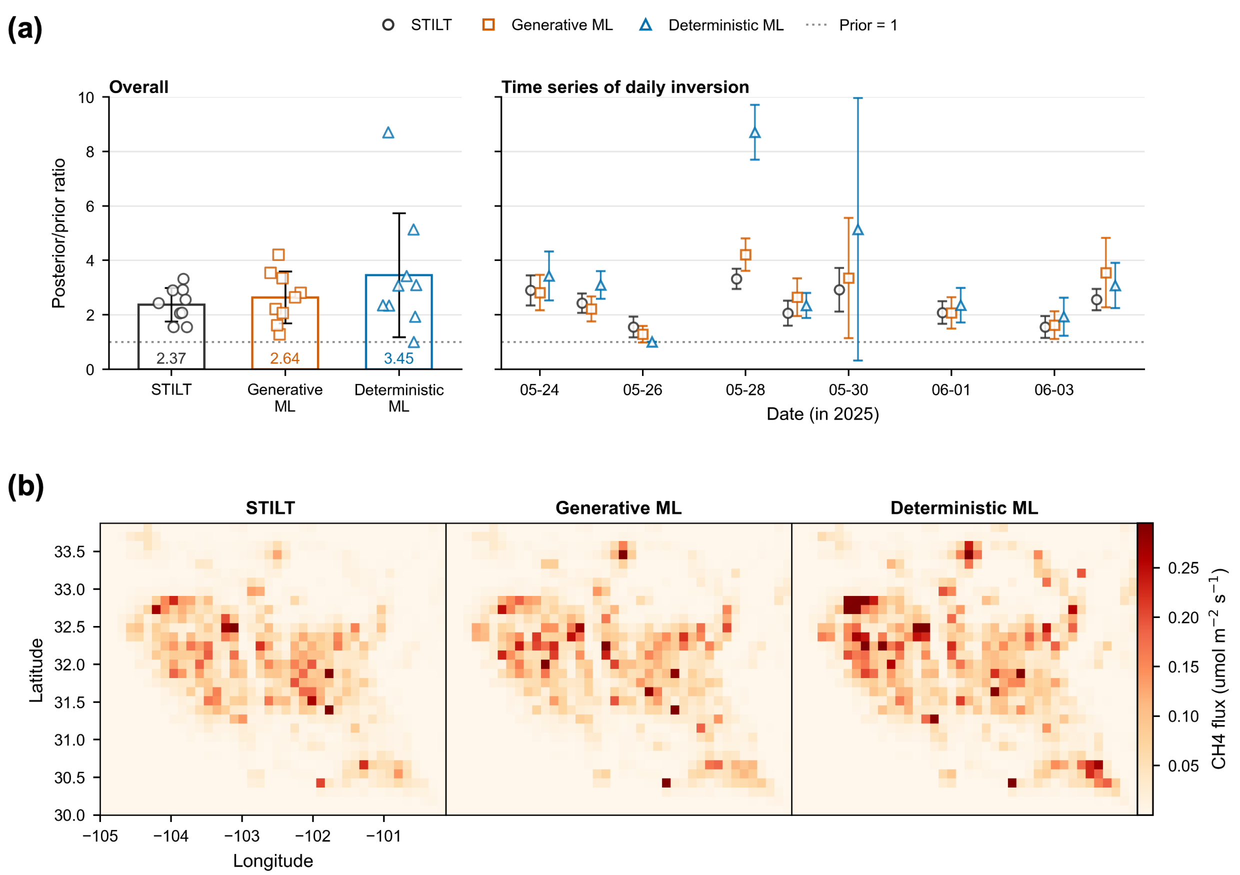


**Figure 6. OSSE methane inverse modeling with three transport operators (STILT, Generative ML, and deterministic ML) under 5 ppb observation noise.** (a) Posterior/prior ratios of total emissions for daily inverse model runs. (b) Multi-day mean posterior $CH_4$ flux maps.

## 3.3 Cross-product transport diagnosis

GenGHG provides a low-cost way to diagnose when GHG transport estimates are meteorologically stable and when they are sensitive to the chosen forcing. We used this capability to test how differences among meteorological products propagate into footprint differences. The comparison includes products from the NOAA family (GFS and FNL) and the ECMWF family (ERA5 and IFS), spanning real-time forecasts (GFS and IFS) and later analysis or reanalysis products (FNL and ERA5).

The case visualizations in Fig. 7 show how product-dependent footprint differences appear at individual receptors. These cases span consistent footprints across all products (case 1), cross-family differences in source-region extent or transport direction with relatively consistent intra-family footprints (cases 2 and 3), and divergent footprints across all products (case 4). These examples also show that the emulator error, i.e., the difference between the GenGHG and STILT footprints within each

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


case (subplots a–b), can be smaller than the footprint differences caused by meteorological forcing (subplots b–e). Product choice can therefore alter both footprint direction and spatial extent.

Figure 8 provides a benchmark-wide statistical view of footprint differences driven by weather products across all test receptors. GenGHG maps the transport characteristics embedded in each forcing field into footprint differences, measured by total mass and pixel-wise PCC. Using GFS-driven STILT footprints as the base reference, the ECMWF-driven predictions exhibit higher footprint total mass, with relative biases of $+9.63\%$ for ERA5 and $+10.14\%$ for IFS. These product-dependent differences exceed the mean GenGHG emulator bias estimated under GFS forcing ($+2.67\%$), indicating that meteorological forcing can be a larger source of transport-pattern variation than the ML approximation itself. GenGHG further enables pairwise comparisons among products by generating footprints for all samples under each forcing field. Intra-family products are most consistent: GFS and FNL have a total mass correlation of $r = 0.92$ and a median pixel-wise PCC of $0.94$ (interquartile range (IQR), $0.90$–$0.98$), and IFS and ERA5 yield $r = 0.89$ and a median PCC of $0.91$ (IQR, $0.81$–$0.96$). Cross-family differences are larger. The strongest contrast occurs between GFS and IFS, two real-time forecast products, with a total mass correlation of $r = 0.78$ and a median pixel-wise PCC of $0.83$ (IQR, $0.68$–$0.92$). These benchmark-wide statistics, together with the case examples, show that meteorological forcing can produce substantial footprint differences, and that GenGHG can diagnose these differences efficiently across products.

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


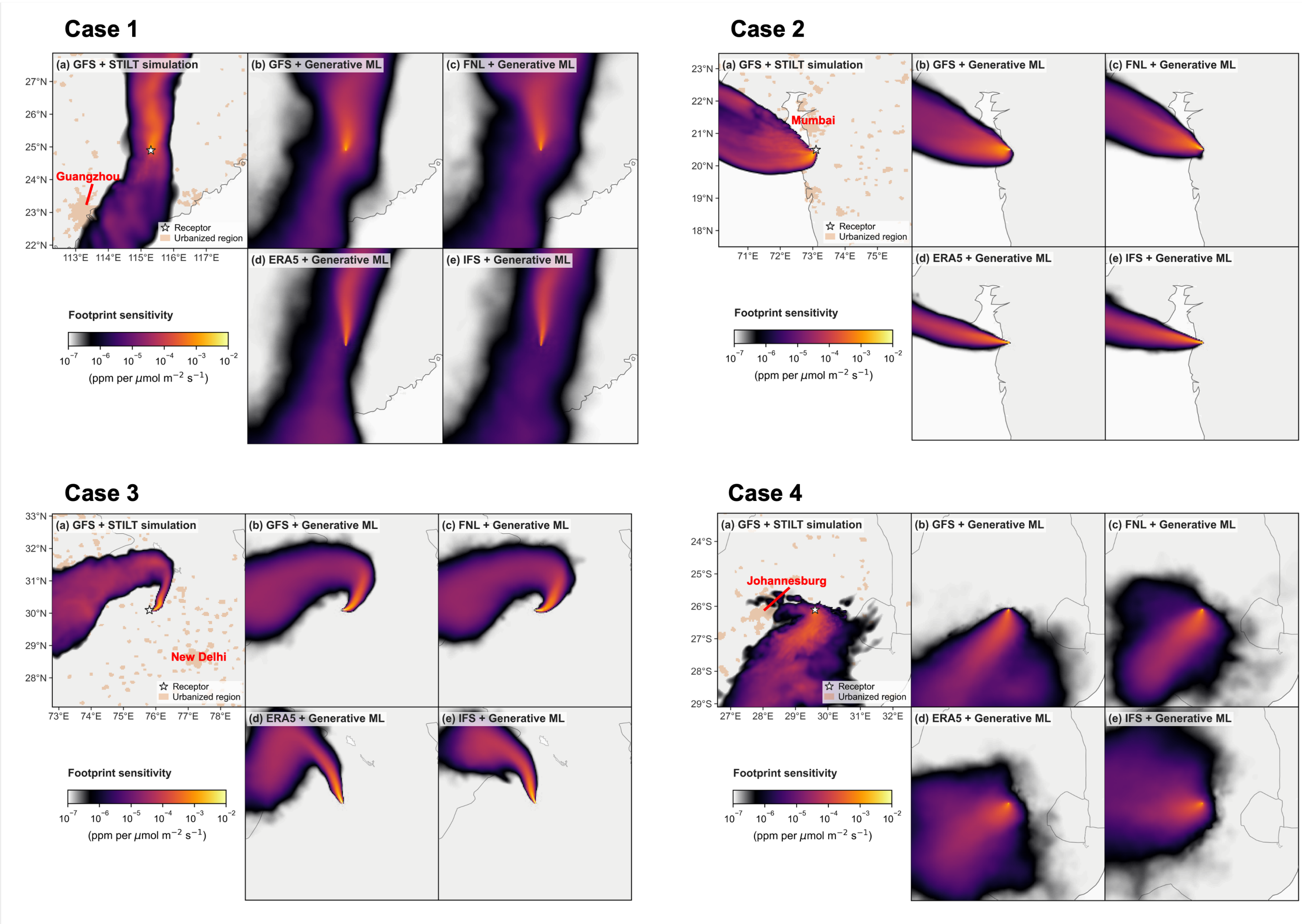


**Figure 7. Cross-product footprint cases.** Four receptor cases compare GFS-driven STILT footprints with GenGHG footprints driven by GFS, FNL, ERA5 and IFS.

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


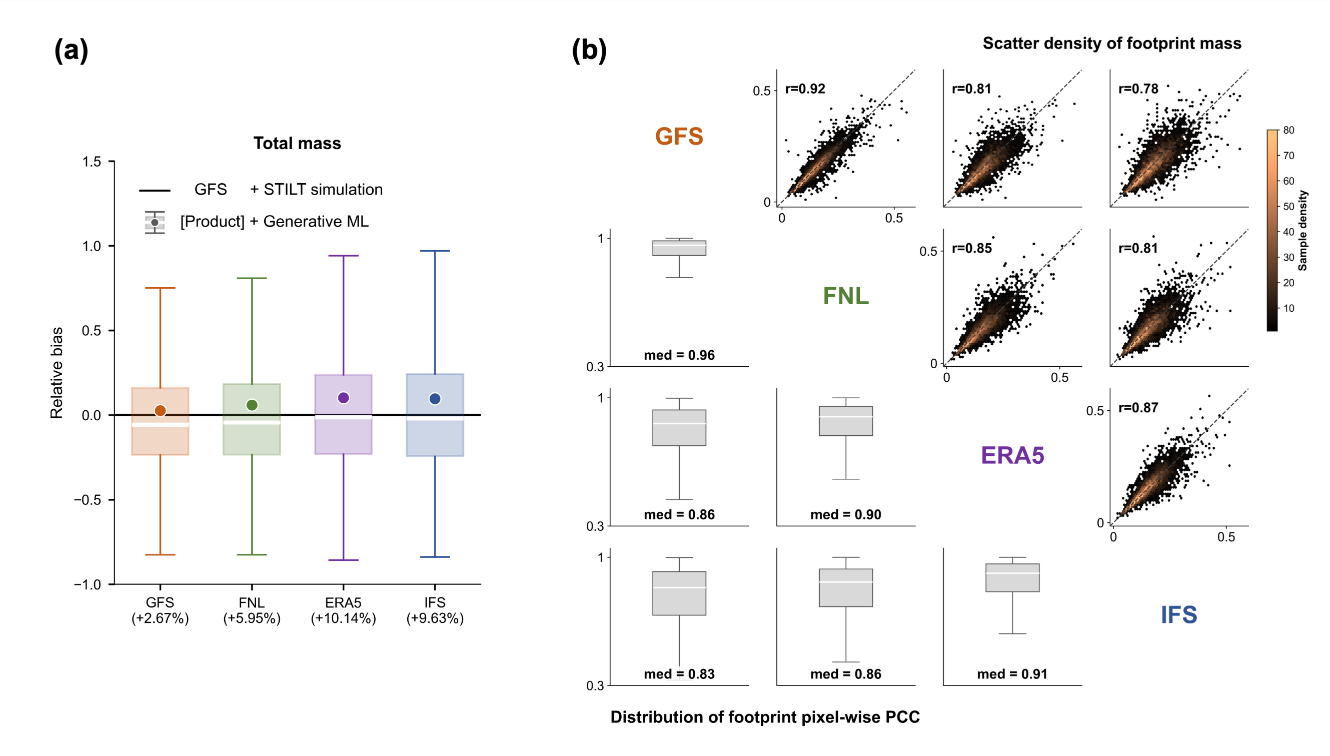


**Figure 8. Cross-product transport sensitivity.** (a) Relative bias in total mass for GenGHG footprints driven by each meteorological product, using GFS-driven STILT footprints as the reference. (b) Pairwise product comparisons: upper-triangle panels show per-sample scatter density of total mass, and lower-triangle panels show per-sample distributions of pixel-wise PCC.

## 4 Discussion and conclusions

Increasing GHG observing density and coverage create new opportunities for top-down monitoring of emission hotspots worldwide. Realizing this opportunity requires transport models that can link surface emissions to atmospheric observations both accurately and rapidly. This study introduces GenGHG as a generative ML model for emulating GHG transport footprints over global urban emission hotspots. GenGHG follows the computational efficiency advantage shared by ML models, while addressing two requirements for inverse modeling use: skilled emulation and operational flexibility.

**Skilled emulation: GenGHG preserves footprint mass, structure, and spatial pattern.** GenGHG reformulates Lagrangian-style footprint emulation as stochastic ensemble prediction rather than point prediction. By representing a family of plausible transport states, the generative model substantially outperforms the deterministic baseline, consistent with the limitations of point-estimate ML emulators reported in previous studies (Fillola et al., 2023; He et al., 2025; Fillola et al., 2026). GenGHG keeps nearly unbiased total mass and structural consistency with the STILT reference (Fig. 3), and preserves the dispersive spatial pattern (Fig. 4). This advantage carries through to the methane inverse modeling case study, where GenGHG produces

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


posterior estimates closer to the STILT reference than deterministic ML (Fig. 6). More broadly, GenGHG demonstrates that Lagrangian transport can be represented as a learnable flow in neural-network latent space (Craig et al., 2026), giving the emulation a stronger physical basis than a direct end-to-end mapping between input and output. Our generative modeling pipeline therefore provides a transferable route for related LPDM frameworks and other trace gases.

**Operational flexibility: GenGHG scales to global emission hotspots and multiple weather products.** GenGHG was developed and evaluated on a benchmark spanning global urban emission hotspots, making it suitable for spatially dense, globally sampled satellite observation settings. Beyond fast footprint generation, its use of $0.25^\circ$ meteorological forcing fields allows footprints to be generated from different publicly-available weather products and used to diagnose forcing-dependent transport differences (Figs. 7 and 8). This capability complements full-physics models such as STILT, for which repeated simulations across many observations and meteorological products can be computationally prohibitive. In operational settings, the ML formulation may provide more stable footprint predictions when full-physics workflows become sensitive to mass errors, as suggested by the stress-test analysis (Fig. 3). We expect GenGHG's operational role to grow as satellite coverage increases and ML-based weather prediction systems provide additional global forcing options (e.g., Sadeghi Tabas et al., 2025; Moldovan et al., 2026).

The present version of GenGHG also points to several directions for further development. Complex topography remains a potential limitation of our benchmark because fewer quality-controlled physical references are available in mountainous regions. Addressing this limitation will require expanding the benchmark with footprint simulations driven by nested, higher-resolution meteorology beyond the current $0.25^\circ$ forcing, and incorporating explicit topographic fields as additional predictors for GenGHG. More broadly, the same framework could be extended from localized emission hotspots to regional or continental transport, and from satellite column receptors to other observing platforms. Finally, beyond using the ensemble mean as the transport estimate, future work should test whether the spread among GenGHG ensemble members can quantify the uncertainty of transport model.

## Appendix A: GenGHG model configuration and training settings

GenGHG is implemented in Python 3.9 and PyTorch 1.10, and is trained in parallel on two NVIDIA RTX 3090 GPUs with CUDA 11.3. GenGHG contains two separately trained modules: the VAE-based subgrid adapter is trained first and then fixed during training of the transformer-based transport module.

The primary transport model consists of input embeddings, a transformer encoder and an output decoder. The footprint state $\boldsymbol{h}_t$, normalized generative time $t$ and meteorological forcing condition $\boldsymbol{c}$ are first embedded into 512 feature channels before entering the transformer module. The transformer contains 6 self-attention blocks, each with 8 attention heads, a hidden width of 512, a feed-forward width of 2048 and a dropout rate of 0.1. It predicts the velocity field used by the flow-matching generation process. The output decoder maps the 512-channel hidden representation to a 16-channel latent velocity field, which is aligned with the VAE latent footprint space and can be decoded back to the original footprint grid. The parameters are optimized with AdamW (Loshchilov and Hutter, 2017) using a batch size of 64. The initial learning rate is $10^{-3}$, with a

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


warm-up during the first 2 epochs and a multi-step learning-rate schedule with a decay factor over epochs 2–200. Training uses mixed precision and gradient clipping with a maximum norm of 0.999. The final checkpoint is selected using validation RMSE from ensemble prediction.

The VAE maps between the native footprint grid and the latent footprint space, serving as an $8\times$ subgrid adapter from the $192 \times 192 \times 1$ footprint grid (height $\times$ width $\times$ channel) to a $24 \times 24 \times 16$ latent grid. It uses 512 hidden channels and a 16-channel latent embedding: the encoder compresses footprints through three downsampling convolutional stages and residual blocks, whereas the decoder reconstructs them through transposed-convolution upsampling. The parameters are trained with AdamW for 60 epochs using a batch size of 256, with a loss combining reconstruction error and a Kullback–Leibler (KL) divergence term weighted by $10^{-6}$. During GenGHG prediction, only the trained decoder of VAE is used to map the latent footprints back to the native grid.

## Appendix B: STILT footprint screening with the dmas diagnostic

We screen each simulated footprint using the STILT "destruction of mass" diagnostic (dmas). The dmas value is a dimensionless particle-level factor for mass correction; its distribution summarizes how strongly particles deviate from mass-conserving transport during the backward simulation. We use robust percentiles of particle-level dmas values to identify anomalous simulations:

$$\begin{aligned}
\text{pass}: &\quad P_{99}(\mathrm{dmas}) < 5, \quad 0.8 < P_{50}(\mathrm{dmas}) < 1.2, \quad P_5(\mathrm{dmas}) > 0.1,\\
\text{badly-fail}: &\quad P_{99}(\mathrm{dmas}) > 30 \quad \text{or} \quad P_{15}(\mathrm{dmas}) = 0,\\
\text{fail}: &\quad \text{otherwise.}
\end{aligned} \tag{B1}$$

Large upper-tail dmas values indicate excessive mass correction and can produce artificial near-surface sensitivity peaks, leading to nonphysical surface trapping and a systematic overestimation of ground-level concentrations; whereas zero lower-tail values indicate severe dilution that can suppress upstream influence

## Appendix C: Terrain classification of benchmark cities

The 60 urban regions were assigned to five terrain-complexity classes (T0–T4) using their geomorphic setting and the terrain descriptors of Amatulli et al. (2018), including elevation context, slope, relief, roughness, topographic position and landform type (Table C1). This classification is used only to stratify benchmark coverage and model performance by terrain context, not as a substitute for site-specific high-resolution geomorphometric analysis.

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


**Table C1.** Terrain classes of the benchmark urban regions.

| Class | Terrain class | Classification basis | Benchmark urban regions |
|---|---|---|---|
| T0 | Coastal lowland | Low-relief coastal, deltaic or reclaimed lowland | Amsterdam; Bangkok; Buenos Aires; Guangzhou; Ho Chi Minh City; Jakarta; Lagos; Manila; Mumbai; New York; Shanghai |
| T1 | Inland plain | Low-relief inland fluvial, lacustrine or glacial plain | Baghdad; Berlin; Cairo; Chicago; Khartoum; Lahore; London; Melbourne; Minneapolis; Moscow; New Delhi; Paris; Shenyang; Toronto; Wuhan; Zhengzhou |
| T2 | Upland or plateau | Rolling upland, plateau or dissected low-hill terrain | Atlanta; Bangalore; Charlotte; Chengdu; Dallas; Johannesburg; Kansas City; Madrid; Milan; Riyadh; Rome; Sao Paulo |
| T3 | Hilly terrain | Hilly, foothill, valley-margin or basin-margin terrain | Amman; Barcelona; Beijing; Durban; Esfahan; Istanbul; Kuala Lumpur; Lyon; Seattle; Seoul; Sydney; Tokyo; Tashkent; Vienna |
| T4 | Mountain | Mountain-front, intermontane or steep coastal-mountain terrain | Cape Town; Los Angeles; Mexico City; Salt Lake City; Taipei; Tehran; Urumqi |

## Appendix D: Bayesian inverse model

For each transport option, posterior fluxes were estimated from the synthetic $XCH_4$ observations using the Bayesian inverse model (Michalak et al., 2004) as

$$\hat{s} = \boldsymbol{s}_{\mathrm{p}} + \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{Q}^{-1}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\left(\boldsymbol{y}^{\mathrm{syn}} - \mathbf{H}\boldsymbol{s}_{\mathrm{p}}\right), \tag{D1}$$

where $\hat{s}$ is the posterior flux estimate, $\boldsymbol{s}_{\mathrm{p}}$ is the prior flux field, $\boldsymbol{y}^{\mathrm{syn}}$ is the synthetic observation vector and $\mathbf{H}$ is the transport operator used in the inverse model; $\mathbf{R}$ is the model–data mismatch covariance and $\mathbf{Q}$ is the prior error covariance. In our design of OSSE, both covariance matrices were estimated from the data using the same maximum-likelihood procedure for all transport options (Michalak et al., 2005). This data-driven treatment avoids imposing fixed error settings and keeps the comparison focused on the transport matrix $\mathbf{H}$.

For each daily inverse model run, the corresponding posterior error covariance is

$$\mathbf{V}_{\hat{s}} = \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{Q}^{-1}\right)^{-1} \tag{D2}$$

which describes the remaining flux uncertainty after combining the prior information with the synthetic observations. In this study, grid-level posterior uncertainty was summarized from $\mathbf{V}_{\hat{s}}$ and aggregated over the source domain to report the uncertainty of each daily total-emission estimate.

The GenGHG v1.0 code, Python package and trained model checkpoints used in this study are archived on Zenodo at https://doi.org/10.5281/zenodo.21017062 (Wang and Wang, 2026). The benchmark dataset used for model training, validation and testing, including the processed meteorological inputs and STILT footprint targets needed to reproduce the reported experiments, is archived on Zenodo at https://doi.org/10.5281/zenodo.21053452 (Wang, 2026). The physical reference footprints were produced with the STILT-R code (Fasoli, 2018), with column pressure weighting adapted from the X-STILT code (Wu, 2023). The original meteorological products were obtained from institutional archives: GFS from the NCAR Geoscience Data Exchange (GDEX) archive available at https://gdex.ucar.edu/datasets/d084001/; FNL from the NCAR GDEX

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


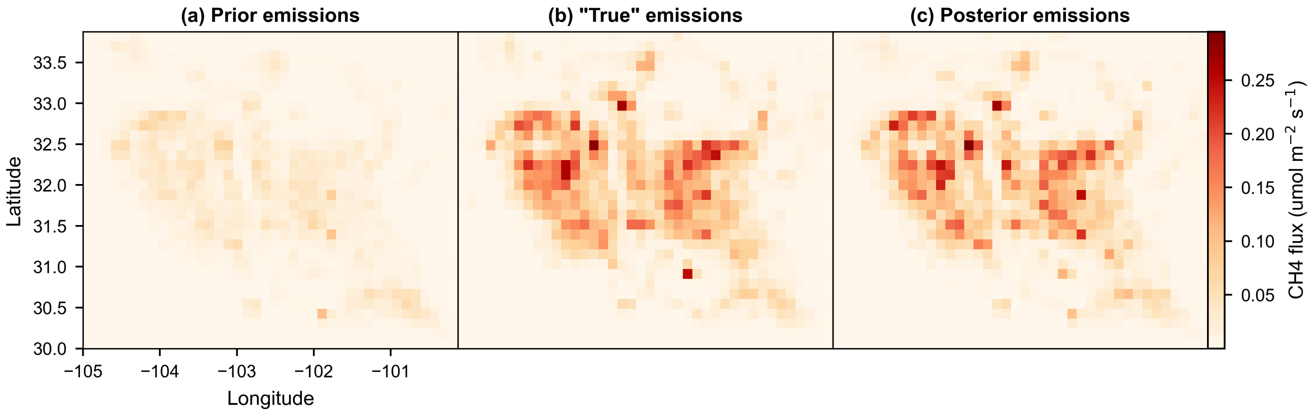


**Figure D1. Noise-free OSSE reference inverse modeling.** (a) Prior emissions from the EPA GHGI inventory. (b) Prescribed true emissions from the measurement-based oil-and-gas methane inventory (Omara et al., 2024). (c) Posterior emissions inferred using STILT as the reference transport operator. Recovery of the main true-emission structure confirms that the synthetic receptor configuration provides effective inverse modeling constraints.

**Table D1. OSSE inverse modeling diagnostics across observation-noise levels.** Values are reported as mean ±standard deviation. Posterior/prior ratios of total emissions are averaged across all daily inverse model runs. Model–observation mismatch values are $XCH_4$ residuals in ppb, summarized over all synthetic observations. The STILT column gives the reference case without transport-operator error.

| Diagnostic | | Noise level (ppb) | Transport operator | | |
|---|---|---|---|---|---|
| | | | STILT | GenGHG | DeterGHG |
| Posterior/prior ratio of total emissions | | 0 | 2.976 ±0.313 | 3.294 ±0.898 | 4.962 ±2.404 |
| | | 5 | 2.372 ±0.615 | 2.638 ±0.950 | 3.452 ±2.278 |
| | | 8 | 2.362 ±0.628 | 2.761 ±1.048 | 3.287 ±2.191 |
| | | 10 | 2.035 ±0.703 | 2.236 ±1.144 | 2.830 ±2.088 |
| Model–observation mismatch (ppb) | Posterior | 0 | -0.000 ±0.002 | -0.223 ±1.337 | -0.396 ±1.520 |
| | | 5 | -0.141 ±4.799 | -0.374 ±5.093 | -0.695 ±5.162 |
| | | 8 | -0.558 ±7.808 | -0.734 ±7.951 | -1.197 ±7.973 |
| | | 10 | -0.480 ±9.638 | -0.654 ±9.808 | -1.009 ±9.765 |
| | Prior | 0 | -2.443 ±3.532 | -2.672 ±4.067 | -2.953 ±4.378 |
| | | 5 | -2.298 ±6.117 | -2.527 ±6.437 | -2.809 ±6.643 |
| | | 8 | -2.719 ±8.714 | -2.948 ±8.949 | -3.230 ±9.100 |
| | | 10 | -2.306 ±10.445 | -2.535 ±10.623 | -2.817 ±10.751 |

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


archive available at https://gdex.ucar.edu/datasets/d083003/; IFS forecasts from the ECMWF open-data archive available at
420 https://registry.opendata.aws/ecmwf-forecasts/; and ERA5 single-level and pressure-level products from the Copernicus Climate Data Store available at https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels and https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels.

*Author contributions.* ZW conceptualized the study and developed the methodology. ZW and SMM designed the validation strategy, and ZW performed the analyses and wrote the original draft. JW contributed to methodology development and validation. HL, ZH, KL, FZ and
425 SMM reviewed and edited the manuscript. FZ acquired funding and supervised the project.

*Competing interests.* The authors declare that they have no conflict of interest.

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


## References

Alet, F., Price, I., El-Kadi, A., Masters, D., Markou, S., Andersson, T. R., Stott, J., Lam, R., Willson, M., Sanchez-Gonzalez, A., and Battaglia, P.: Skillful Joint Probabilistic Weather Forecasting from Marginals, https://doi.org/10.48550/arXiv.2506.10772, 2025.

Amatulli, G., Domisch, S., Tuanmu, M.-N., Parmentier, B., Ranipeta, A., Malczyk, J., and Jetz, W.: A suite of global, cross-scale topographic variables for environmental and biodiversity modeling, Scientific data, 5, 180 040, 2018.

Balasus, N., Jacob, D. J., Lorente, A., Maasakkers, J. D., Parker, R. J., Boesch, H., Chen, Z., Kelp, M. M., Nesser, H., and Varon, D. J.: A Blended TROPOMI+GOSAT Satellite Data Product for Atmospheric Methane Using Machine Learning to Correct Retrieval Biases, Atmospheric Measurement Techniques, 16, 3787–3807, https://doi.org/10.5194/amt-16-3787-2023, 2023.

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q., Liu, H. Y., Mickley, L. J., and Schultz, M. G.: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, Journal of Geophysical Research: Atmospheres, 106, 23 073–23 095, 2001.

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

Copernicus Climate Change Service, Climate Data Store: ERA5 hourly data on pressure levels from 1940 to present, https://doi.org/10.24381/cds.bd0915c6, accessed: 01-Feb-2026, 2023a.

Copernicus Climate Change Service, Climate Data Store: ERA5 hourly data on single levels from 1940 to present, https://doi.org/10.24381/cds.adbb2d47, accessed: 01-Feb-2026, 2023b.

Couairon, G., Singh, R., Charantonis, A., Lessig, C., and Monteleoni, C.: ArchesWeatherGen: Skillful and Compute-Efficient Probabilistic Weather Forecasting with Machine Learning, Science Advances, 12, eadx2372, https://doi.org/10.1126/sciadv.adx2372, 2026.

Craig, G., Selz, T., Beylich, M., and Tempest, K. I.: The Physics of AI Weather Models, https://doi.org/10.48550/arXiv.2605.23778, 2026.

Cusworth, D. H., Bloom, A. A., Ma, S., Miller, C. E., Bowman, K., Yin, Y., Maasakkers, J. D., Zhang, Y., Scarpelli, T. R., Qu, Z., Jacob, D. J., and Worden, J. R.: A Bayesian Framework for Deriving Sector-Based Methane Emissions from Top-down Fluxes, Communications Earth & Environment, 2, 242, https://doi.org/10.1038/s43247-021-00312-6, 2021.

Dadheech, N. and Turner, A. J.: Simulating Out-of-Sample Atmospheric Transport to Enable Flux Inversions, Atmospheric Chemistry and Physics, 26, 427–441, https://doi.org/10.5194/acp-26-427-2026, 2026.

Dao, Q., Phung, H., Nguyen, B., and Tran, A.: Flow matching in latent space, arXiv preprint arXiv:2307.08698, 2023.

Deng, A., Lauvaux, T., Davis, K. J., Gaudet, B. J., Miles, N., Richardson, S. J., Wu, K., Sarmiento, D. P., Hardesty, R. M., Bonin, T. A., Brewer, W. A., and Gurney, K. R.: Toward Reduced Transport Errors in a High Resolution Urban CO2 Inversion System, Elementa: Science of the Anthropocene, 5, 20, https://doi.org/10.1525/elementa.133, 2017.

Dumont Le Brazidec, J., Vanderbecken, P., Farchi, A., Bocquet, M., Lian, J., Broquet, G., Kuhlmann, G., Danjou, A., and Lauvaux, T.: Segmentation of $XCO_2$ Images with Deep Learning: Application to Synthetic Plumes from Cities and Power Plants, Geoscientific Model Development, 16, 3997–4016, https://doi.org/10.5194/gmd-16-3997-2023, 2023.

Duren, R. M., Thorpe, A. K., Foster, K. T., Rafiq, T., Hopkins, F. M., Yadav, V., Bue, B. D., Thompson, D. R., Conley, S., Colombi, N. K., Frankenberg, C., McCubbin, I. B., Eastwood, M. L., Falk, M., Herner, J. D., Croes, B. E., Green, R. O., and Miller, C. E.: California's Methane Super-Emitters, Nature, 575, 180–184, https://doi.org/10.1038/s41586-019-1720-3, 2019.

Eldering, A., Wennberg, P., Crisp, D., Schimel, D., Gunson, M., Chatterjee, A., Liu, J., Schwandner, F., Sun, Y., O'dell, C., et al.: The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes, Science, 358, eaam5745, 2017.

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


Estrada, L. A., Varon, D. J., Sulprizio, M., Nesser, H., Chen, Z., Balasus, N., Hancock, S. E., He, M., East, J. D., Mooring, T. A., Oort Alonso, A., Maasakkers, J. D., Aben, I., Baray, S., Bowman, K. W., Worden, J. R., Cardoso-Saldaña, F. J., Reidy, E., and Jacob, D. J.: Integrated Methane Inversion (IMI) 2.0: An Improved Research and Stakeholder Tool for Monitoring Total Methane Emissions with High Resolution Worldwide Using TROPOMI Satellite Observations, https://doi.org/10.5194/egusphere-2024-2700, 2024.

European Centre for Medium-Range Weather Forecasts: ECMWF Open Data: IFS forecast data, https://doi.org/10.21957/open-data, accessed: 01-Feb-2026. Data are licensed under CC BY 4.0, 2026.

Fasoli, B.: uataq/stilt: Geoscientific Model Development 2018, https://doi.org/10.5281/zenodo.1238047, 2018.

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., 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, https://doi.org/10.5194/gmd-19-1893-2026, 2026.

Ganshin, A., Oda, T., Saito, M., Maksyutov, S., Valsala, V., Andres, R. J., Fisher, R. E., Lowry, D., Lukyanov, A., Matsueda, H., Nisbet, E. G., Rigby, M., Sawa, Y., Toumi, R., Tsuboi, K., Varlagin, A., and Zhuravlev, R.: A Global Coupled Eulerian-Lagrangian Model and 1 × 1 Km $CO_2$ Surface Flux Dataset for High-Resolution Atmospheric $CO_2$ Transport Simulations, Geoscientific Model Development, 5, 231–243, https://doi.org/10.5194/gmd-5-231-2012, 2012.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward Constraining Regional-Scale Fluxes of CO2 with Atmospheric Observations over a Continent: 2. Analysis of COBRA Data Using a Receptor-Oriented Framework, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2003JD003770, 2003a.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward Constraining Regional-Scale Fluxes of CO2 with Atmospheric Observations over a Continent: 2. Analysis of COBRA Data Using a Receptor-Oriented Framework, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2003JD003770, 2003b.

Gerbig, C., Körner, S., and Lin, J. C.: Vertical Mixing in Atmospheric Tracer Transport Models: Error Characterization and Propagation, Atmospheric Chemistry and Physics, 8, 591–602, https://doi.org/10.5194/acp-8-591-2008, 2008.

He, T.-L., Dadheech, N., Thompson, T. M., and Turner, A. J.: FootNet v1.0: Development of a Machine Learning Emulator of Atmospheric Transport, Geoscientific Model Development, 18, 1661–1671, https://doi.org/10.5194/gmd-18-1661-2025, 2025.

He, W., Van Der Velde, I. R., Andrews, A. E., Sweeney, C., Miller, J., Tans, P., Van Der Laan-Luijkx, I. T., Nehrkorn, T., Mountain, M., Ju, W., Peters, W., and Chen, H.: CTDAS-Lagrange v1.0: A High-Resolution Data Assimilation System for Regional Carbon Dioxide Observations, Geoscientific Model Development, 11, 3515–3536, https://doi.org/10.5194/gmd-11-3515-2018, 2018.

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., et al.: Quantifying methane emissions from the global scale down to point sources using satellite observations of atmospheric methane, Atmospheric Chemistry and Physics, 22, 9617–9646, 2022.

Janssens-Maenhout, G., Pinty, B., Dowell, M., Zunker, H., Andersson, E., Balsamo, G., Bézy, J.-L., Brunhes, T., Bösch, H., Bojkov, B., et al.: Toward an operational anthropogenic CO 2 emissions monitoring and verification support capacity, Bulletin of the American Meteorological Society, 101, E1439–E1451, 2020.

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.

Karion, A., Callahan, W., Stock, M., Prinzivalli, S., Verhulst, K. R., Kim, J., Salameh, P. K., Lopez-Coto, I., and Whetstone, J.: Greenhouse gas observations from the Northeast Corridor tower network, Earth System Science Data, 12, 699–717, 2020.

Kiel, M., Eldering, A., Roten, D. D., Lin, J. C., Feng, S., Lei, R., Lauvaux, T., Oda, T., Roehl, C. M., Blavier, J.-F., et al.: Urban-focused satellite CO2 observations from the Orbiting Carbon Observatory-3: A first look at the Los Angeles megacity, Remote Sensing of Environment, 258, 112 314, 2021.

Kingma, D. P. and Welling, M.: Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114, 2013.

Krol, M., Houweling, S., Bregman, B., Van den Broek, M., Segers, A., Van Velthoven, P., Peters, W., Dentener, F., and Bergamaschi, P.: The two-way nested global chemistry-transport zoom model TM5: algorithm and applications, Atmospheric Chemistry and Physics, 5, 417–432, 2005.

Lang, S., Alexe, M., Clare, M. C. A., Roberts, C., Adewoyin, R., Ben Bouallègue, Z., Chantry, M., Dramsch, J., Dueben, P. D., Hahner, S., Maciel, P., Prieto-Nemesio, A., O'Brien, C., Pinault, F., Polster, J., Raoult, B., Tietsche, S., and Leutbecher, M.: AIFS-CRPS: Ensemble Forecasting Using a Model Trained with a Loss Function Based on the Continuous Ranked Probability Score, npj Artificial Intelligence, 2, 18, https://doi.org/10.1038/s44387-026-00073-7, 2026.

Lauvaux, T., Miles, N. L., Deng, A., Richardson, S. J., Cambaliza, M. O., Davis, K. J., Gaudet, B., Gurney, K. R., Huang, J., O'Keefe, D., et al.: High-resolution atmospheric inversion of urban CO2 emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX), Journal of Geophysical Research: Atmospheres, 121, 5213–5236, 2016.

Lauvaux, T., Giron, C., Mazzolini, M., d'Aspremont, A., Duren, R., Cusworth, D., Shindell, D., and Ciais, P.: Global assessment of oil and gas methane ultra-emitters, Science, 375, 557–561, 2022.

Lin, J. C. and Gerbig, C.: Accounting for the Effect of Transport Errors on Tracer Inversions, Geophysical Research Letters, 32, 2004GL021 127, https://doi.org/10.1029/2004GL021127, 2005.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A Near-Field Tool for Simulating the Upstream Influence of Atmospheric Observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) Model, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002JD003161, 2003.

Lipman, Y., Chen, R. T., Ben-Hamu, H., Nickel, M., and Le, M.: Flow matching for generative modeling, arXiv preprint arXiv:2210.02747, 2022.

Liu, H., Ishizawa, M., Vogel, F. R., Zhang, Z., Poulter, B., Feng, L., Chen, A., Gagné-Landmann, A. L., Huntzinger, D. N., Melton, J. R., et al.: Methane fluxes from Arctic & boreal North America: comparisons between process-based estimates and atmospheric observations, Atmospheric Chemistry and Physics, 26, 1229–1247, 2026.

Loshchilov, I. and Hutter, F.: Decoupled weight decay regularization, arXiv preprint arXiv:1711.05101, 2017.

Maasakkers, J. D., Jacob, D. J., Sulprizio, M. P., Turner, A. J., Weitz, M., Wirth, T., Hight, C., DeFigueiredo, M., Desai, M., Schmeltz, R., Hockstad, L., Bloom, A. A., Bowman, K. W., Jeong, S., and Fischer, M. L.: Gridded National Inventory of U.S. Methane Emissions, Environmental Science & Technology, 50, 13 123–13 133, https://doi.org/10.1021/acs.est.6b02878, 2016.

Michalak, A. M., Bruhwiler, L., and Tans, P. P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, Journal of Geophysical Research: Atmospheres, 109, 2004.

Michalak, A. M., Hirsch, A., Bruhwiler, L., Gurney, K. R., Peters, W., and Tans, P. P.: Maximum Likelihood Estimation of Covariance Parameters for Bayesian Atmospheric Trace Gas Surface Flux Inversions, Journal of Geophysical Research: Atmospheres, 110, https://doi.org/10.1029/2005JD005970, 2005.

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


Moldovan, G., Pinnington, E., Prieto Nemesio, A., Lang, S., Ben Bouallègue, Z., Dramsch, J., Alexe, M., Santa Cruz, M., Hahner, S., Cook, H., et al.: AIFS Single 1.1. 0: an update to ECMWF's machine-learned weather forecast model AIFS, Geoscientific Model Development, 19, 4703–4724, 2026.

Moran, D., Kanemoto, K., Jiborn, M., Wood, R., Többen, J., and Seto, K. C.: Carbon footprints of 13 000 cities, Environmental Research Letters, 13, 064 041, 2018.

Munassar, S., Monteil, G., Scholze, M., Karstens, U., Rödenbeck, C., Koch, F.-T., Totsche, K. U., and Gerbig, C.: Why Do Inverse Models Disagree? A Case Study with Two European $CO_2$ Inversions, Atmospheric Chemistry and Physics, 23, 2813–2828, https://doi.org/10.5194/acp-23-2813-2023, 2023.

Nassar, R., Hill, T. G., McLinden, C. A., Wunch, D., Jones, D. B., and Crisp, D.: Quantifying CO2 emissions from individual power plants from space, Geophysical Research Letters, 44, 10–045, 2017.

Nassar, R., Mastrogiacomo, J.-P., Bateman-Hemphill, W., McCracken, C., MacDonald, C. G., Hill, T., O'Dell, C. W., Kiel, M., and Crisp, D.: Advances in quantifying power plant CO2 emissions with OCO-2, Remote Sensing of Environment, 264, 112 579, 2021.

Nehrkorn, T., Eluszkiewicz, J., Wofsy, S. C., Lin, J. C., Gerbig, C., Longo, M., and Freitas, S.: Coupled Weather Research and Forecasting–Stochastic Time-Inverted Lagrangian Transport (WRF–STILT) Model, Meteorology and Atmospheric Physics, 107, 51–64, https://doi.org/10.1007/s00703-010-0068-x, 2010.

Nichol, A. Q. and Dhariwal, P.: Improved denoising diffusion probabilistic models, in: International conference on machine learning, pp. 8162–8171, PMLR, 2021.

NOAA NCEP: NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids, https://doi.org/10.5065/D65Q4T4Z, national Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce. Accessed: 01-Feb-2026, 2015a.

NOAA NCEP: NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive, https://doi.org/10.5065/D65D8PWK, national Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce. Accessed: 01-Aug-2025, 2015b.

Ohyama, H., Frey, M. M., Morino, I., Shiomi, K., Nishihashi, M., Miyauchi, T., Yamada, H., Saito, M., Wakasa, M., Blumenstock, T., et al.: Anthropogenic CO 2 emission estimates in the Tokyo metropolitan area from ground-based CO 2 column observations, Atmospheric Chemistry and Physics, 23, 15 097–15 119, 2023.

Omara, M., Himmelberger, A., MacKay, K., Williams, J. P., Benmergui, J., Sargent, M., Wofsy, S. C., and Gautam, R.: Constructing a Measurement-Based Spatially Explicit Inventory of US Oil and Gas Methane Emissions (2021), Earth System Science Data, 16, 3973–3991, https://doi.org/10.5194/essd-16-3973-2024, 2024.

Ramaswami, A., Tong, K., Canadell, J. G., Jackson, R. B., Stokes, E., Dhakal, S., Finch, M., Jittrapirom, P., Singh, N., Yamagata, Y., et al.: Carbon analytics for net-zero emissions sustainable cities, Nature Sustainability, 4, 460–463, 2021.

Reuter, M., Buchwitz, M., Hilker, M., Heymann, J., Schneising, O., Pillai, D., Bovensmann, H., Burrows, J. P., Bösch, H., Parker, R., Butz, A., Hasekamp, O., O'Dell, C. W., Yoshida, Y., Gerbig, C., Nehrkorn, T., Deutscher, N. M., Warneke, T., Notholt, J., Hase, F., Kivi, R., Sussmann, R., Machida, T., Matsueda, H., and Sawa, Y.: Satellite-Inferred European Carbon Sink Larger than Expected, Atmospheric Chemistry and Physics, 14, 13 739–13 753, https://doi.org/10.5194/acp-14-13739-2014, 2014.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B.: High-resolution image synthesis with latent diffusion models, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10 684–10 695, 2022.

Sadeghi Tabas, S., Wang, J., Lei, W., Row, M., Zhang, Z., Zhu, L., Peng, J., and Carley, J. R.: GFS-powered machine learning weather prediction: A comparative study on training GraphCast with NOAA's GDAS data for global weather forecasts, 2025.

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


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, https://doi.org/10.1175/BAMS-D-14-00110.1, 2015.

Stohl, A., Hittenberger, M., and Wotawa, G.: Validation of the Lagrangian particle dispersion model FLEXPART against large-scale tracer experiment data, Atmospheric Environment, 32, 4245–4264, 1998.

Tanimoto, H., Matsunaga, T., Someya, Y., Fujinawa, T., Ohyama, H., Morino, I., Yashiro, H., Sugita, T., Inomata, S., Müller, A., Saeki, T., Yoshida, Y., Niwa, Y., Saito, M., Noda, H., Yamashita, Y., Ikeda, K., Saigusa, N., Machida, T., Frey, M. M., Lim, H., Srivastava, P., Jin, Y., Shimizu, A., Nishizawa, T., Kanaya, Y., Sekiya, T., Patra, P., Takigawa, M., Bisht, J., Kasai, Y., and Sato, T. O.: The Greenhouse Gas Observation Mission with Global Observing SATellite for Greenhouse Gases and Water Cycle (GOSAT-GW): Objectives, Conceptual Framework and Scientific Contributions, Progress in Earth and Planetary Science, 12, 8, https://doi.org/10.1186/s40645-025-00684-9, 2025.

Turner, A. J., Jacob, D. J., Benmergui, J., Brandman, J., White, L., and Randles, C. A.: Assessing the capability of different satellite observing configurations to resolve the distribution of methane emissions at kilometer scales, Atmospheric Chemistry and Physics, 18, 8265–8278, 2018.

Upton, S., Reichstein, M., Peters, W., Botía, S., Nelson, J. A., Walther, S., Jung, M., Gans, F., Haszpra, L., and Bastos, A.: Constraining a Data-Driven $CO_2$ Flux Model by Ecosystem and Atmospheric Observations Using Atmospheric Transport, Atmospheric Chemistry and Physics, 26, 2561–2595, https://doi.org/10.5194/acp-26-2561-2026, 2026.

Varon, D. J., Jacob, D. J., Hmiel, B., Gautam, R., Lyon, D. R., Omara, M., Sulprizio, M., Shen, L., Pendergrass, D., Nesser, H., Qu, Z., Barkley, Z. R., Miles, N. L., Richardson, S. J., Davis, K. J., Pandey, S., Lu, X., Lorente, A., Borsdorff, T., Maasakkers, J. D., and Aben, I.: Continuous Weekly Monitoring of Methane Emissions from the Permian Basin by Inversion of TROPOMI Satellite Observations, Atmospheric Chemistry and Physics, 23, 7503–7520, https://doi.org/10.5194/acp-23-7503-2023, 2023.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I.: Attention is all you need, Advances in neural information processing systems, 30, 2017.

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.

Wang, Z.: Benchmark dataset for GenGHG v1.0: generative emulation of greenhouse-gas transport footprints, https://doi.org/10.5281/zenodo.21053452, dataset, 2026.

Wang, Z. and Wang, J.: Code and model checkpoints for GenGHG v1.0: generative emulation of greenhouse-gas transport footprints, https://doi.org/10.5281/zenodo.21017062, computer software, 2026.

Wang, Z., Zhang, F., Wang, J., and Cao, L.: Leveraging Wide Snapshot XCO2 Pre-Training to Estimate Urban Fossil Fuel CO2 Emissions from Space, Remote Sensing of Environment, 335, 115 260, https://doi.org/10.1016/j.rse.2026.115260, 2026.

Wilmot, T. Y., Lin, J. C., Wu, D., Oda, T., and Kort, E. A.: Toward a Satellite-Based Monitoring System for Urban CO2 Emissions in Support of Global Collective Climate Mitigation Actions, Environ. Res. Lett., 2024.

Wu, D.: STILT-NOx v1 (for GMD submission), https://doi.org/10.5281/zenodo.8057850, 2023.

Wu, D., Lin, J. C., Fasoli, B., Oda, T., Ye, X., Lauvaux, T., Yang, E. G., and Kort, E. A.: A Lagrangian Approach towards Extracting Signals of Urban CO2; Emissions from Satellite Observations of Atmospheric Column CO2(XCO2): X-Stochastic Time-Inverted Lagrangian Transport Model ("X-STILT V1"), Geoscientific Model Development, 11, 4843–4871, https://doi.org/10.5194/gmd-11-4843-2018, 2018.

https://doi.org/10.5194/egusphere-2026-4304
Preprint. Discussion started: 4 August 2026


Xu, Y., Hu, Y., Gao, S., Zhu, Q., and Zhang, F.: Multimodal flow matching for large-scale human mobility flow generation using satellite imagery and social sensing data, ISPRS Journal of Photogrammetry and Remote Sensing, 239, 291–308, 2026.

Zhang, J., Chen, J., Wu, K., and Tang, H.: Analyzing the Inversion Performance of a Permanent Urban Column GHG Network: An OSSE Perspective, Earth and Space Science, 12, e2024EA004 175, https://doi.org/10.1029/2024EA004175, 2025.