https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

Physically Structured Target Design Improves the Robustness and Transferability of Neural Network Methane Retrievals

Rongjin Song[1,2], Zhaonan Cai[1,2], Yi Liu[1,2], Dongxu Yang[3], Lin Wu[1], Longfei Tian[4], Denghui Hu[4], Guohua Liu[4]

[1]State Key Laboratory of Atmospheric Environment and Extreme Meteorology, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China.

[2]University of Chinese Academy of Sciences, Beijing 100049, China.

[3]Laboratory of Middle Atmosphere and Global Environment Observation, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China.

[4]Innovation Academy for Microsatellites, Chinese Academy of Sciences, Shanghai 201306, China.

Correspondence to: Zhaonan Cai (caizhaonan@mail.iap.ac.cn)

Abstract

Artificial neural networks (ANNs) offer a computationally efficient alternative to conventional physics-based satellite retrievals of atmospheric methane (CH_4). However, the impact of target design on retrieval generalization and robustness remains unclear. Here, we develop two ANN-based column-averaged dry-air mole fraction of methane (XCH_4) retrieval frameworks from GOSAT-2 observations. The models are trained with full-physics (ANN-FP) or proxy (ANN-Proxy) retrieval products and evaluated against 2021–2022 observations and TCCON measurements. While both ANNs reproduce their training targets during training, ANN-Proxy is more stable out of sample, whereas ANN-FP shows increasing bias and drift. Mechanistic analyses using feature attribution, local gradient geometry, and environmentally conditioned perturbation modes indicate that ANN-Proxy concentrates over 97% of its attribution within CH_4 and carbon dioxide (CO_2) absorption windows and maintains a coherent sensitivity field (mean gradient-direction similarity 0.96 versus 0.72 for ANN-FP). Furthermore, ANN-Proxy sensitivity directions are less aligned with aerosol- and albedo-induced perturbation modes, indicating reduced environmental coupling. The proxy-oriented target shapes the learned mapping to enhance robustness and transferability. These results suggest that physically structured proxy targets provide a practical strategy for rapid, stable, and operationally deployable neural-network methane retrievals in future carbon-monitoring satellite missions.

1 Introduction

Satellite remote sensing provides a practical means for continuous global monitoring of XCH4. Spaceborne shortwave infrared (SWIR) spectrometers, including SCIAMACHY (Buchwitz et al., 2005), GOSAT (Kuze et al., 2009), GOSAT-2 (Imasu et al., 2023), and TROPOMI (Hu et al., 2018), retrieve XCH_4 from backscattered and reflected solar irradiance. Accurate XCH4 retrieval remains challenging, however, because uncertainties in atmospheric scattering and surface reflectance variability can introduce substantial systematic errors, particularly in FP retrievals (Aben et al., 2007; O'Dell et al., 2012).

To mitigate scattering-related errors, proxy retrieval methods estimate XCH_4 indirectly by scaling a retrieved CH_4 column with a simultaneously retrieved CO_2 column, assuming that both gases are affected similarly by light-path perturbations (Feng et al., 2017; Morino et al., 2011; Schepers et al., 2012). This ratio-based approach partially cancels light-path errors and has therefore been widely used in CH_4 emission studies (Monteil et al., 2013; Pandey et al., 2019; Schneising et al., 2014). However, proxy retrievals depend on accurate CO_2 reference fields and assume relatively smooth CO_2 variability, which may limit their applicability in regions with strong CO_2 gradients (Feng et al., 2017; Jacob et al., 2016).

The emergence of next-generation carbon-monitoring satellites makes rapid and scalable methane retrieval increasingly important. These missions will produce massive data volumes and increasingly require near-real-time or onboard processing capability, especially for rapid identification of localized CH_4 point-source emissions. Conventional physics-based retrieval algorithms are computationally expensive and often rely on auxiliary external information, limiting their suitability for fast operational processing and onboard implementation. ANNs provide an efficient alternative by learning mappings from radiances to geophysical variables and enabling fast inference after

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

offline training. Previous studies have demonstrated the potential of ANN-based retrievals for XCO_2 and XCH_4 (David et al., 2021; Reuter et al., 2025; Xie et al., 2023). However, it remains unclear how the choice of training target influences model robustness, transferability, and physical consistency (Ma et al., 2026).

From a machine learning (ML) perspective, the two targets correspond to different regression problems: FP retrieval requires learning complex scattering-induced nonlinearities, whereas proxy ratios exploit shared error structures between CH_4 and CO_2 that simplify the mapping (LeCun et al., 2015).

Existing studies have largely focused on network architectures and training strategies, whereas the role of supervised target definition has received comparatively little attention. We hypothesize that physically structured targets modify the effective geometry of the retrieval problem by suppressing environmentally induced variability that is difficult for data-driven models to disentangle, thereby improving the learnability and robustness of the inverse mapping. To test this hypothesis, we compare ANN retrievals trained with FP and proxy XCH_4 targets and investigate their internal representations through independent TCCON validation, feature-attribution analysis, perturbation experiments, and gradient-space diagnostics, aiming to establish a mechanistic connection between target design and retrieval generalization.

2 Data

We use GOSAT-2 Level 1B (L1B) radiances as inputs to the ANN models. The Level 1B product provides calibrated spectra and geolocation information and serves as the observational basis for both FP and proxy CH_4 retrievals (Suto et al., 2021; Yoshida et al., 2011).

As supervised targets, we use the official GOSAT-2 Level 2 (L2) XCH_4 products, including both FP and proxy retrievals (version V02.20). The data were obtained from the GOSAT-2 Product Archive (https://prdct.gosat-2.nies.go.jp/). The FP retrieval simultaneously estimates gas concentrations and scattering parameters through iterative radiative transfer inversion (Butz et al., 2011; Yoshida et al., 2011). The proxy XCH_4 product is obtained by scaling the retrieved CH_4/CO_2 column ratio (hereafter denoted as R) with a modeled XCO_2 field from global transport simulations (Parker et al., 2020).

To ensure strict comparability between the two retrieval frameworks, FP and proxy Level 2 products are matched using the same soundingUniqueID. Only soundings available in both products are retained, resulting in a unified data set with the same sample population and identical L1B radiance inputs. This design eliminates sampling-induced differences and enables a direct performance comparison between ANN-FP and ANN-Proxy.

Independent validation is performed with ground-based observations from the TCCON, which provides accurate and stable column measurements of XCH_4 and XCO_2 (Deutscher et al., 2010; Wunch et al., 2011). GOSAT-2 soundings are collocated with each TCCON site within ±8 hr, ±5° longitude, and ±3° latitude. This spatiotemporal matching strategy balances sample size and representativeness while reducing sampling discrepancies between satellite footprints and ground-based measurements (Liang et al., 2017). All matched samples are screened using the Level 2

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

quality flags. The resulting collocated data set is used to assess retrieval accuracy and compare the performance of the FP, proxy, and ANN-based retrievals.

To illustrate the spatial coverage of the training and prediction data sets, Figure 1 presents the global distribution of valid GOSAT-2 soundings used in this study. The training data consist of all matched FP–proxy retrieval pairs acquired during 2020 (Figure 1a), while independent predictions are generated for observations collected during 2021–2022 (Figure 1b). The two data sets exhibit similar global sampling patterns, reflecting the stable observing strategy of GOSAT-2. Dense sampling is observed over northern Africa and other low-cloud regions, whereas coverage is comparatively sparse in persistently cloudy tropical areas and high-latitude regions. The locations of the 12 TCCON stations used for validation are shown as black markers. Their broad geographic distribution enables evaluation of model performance under diverse atmospheric and surface conditions. The broadly consistent spatial distributions between the training and prediction periods help minimize potential biases arising from geographic sampling differences.

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

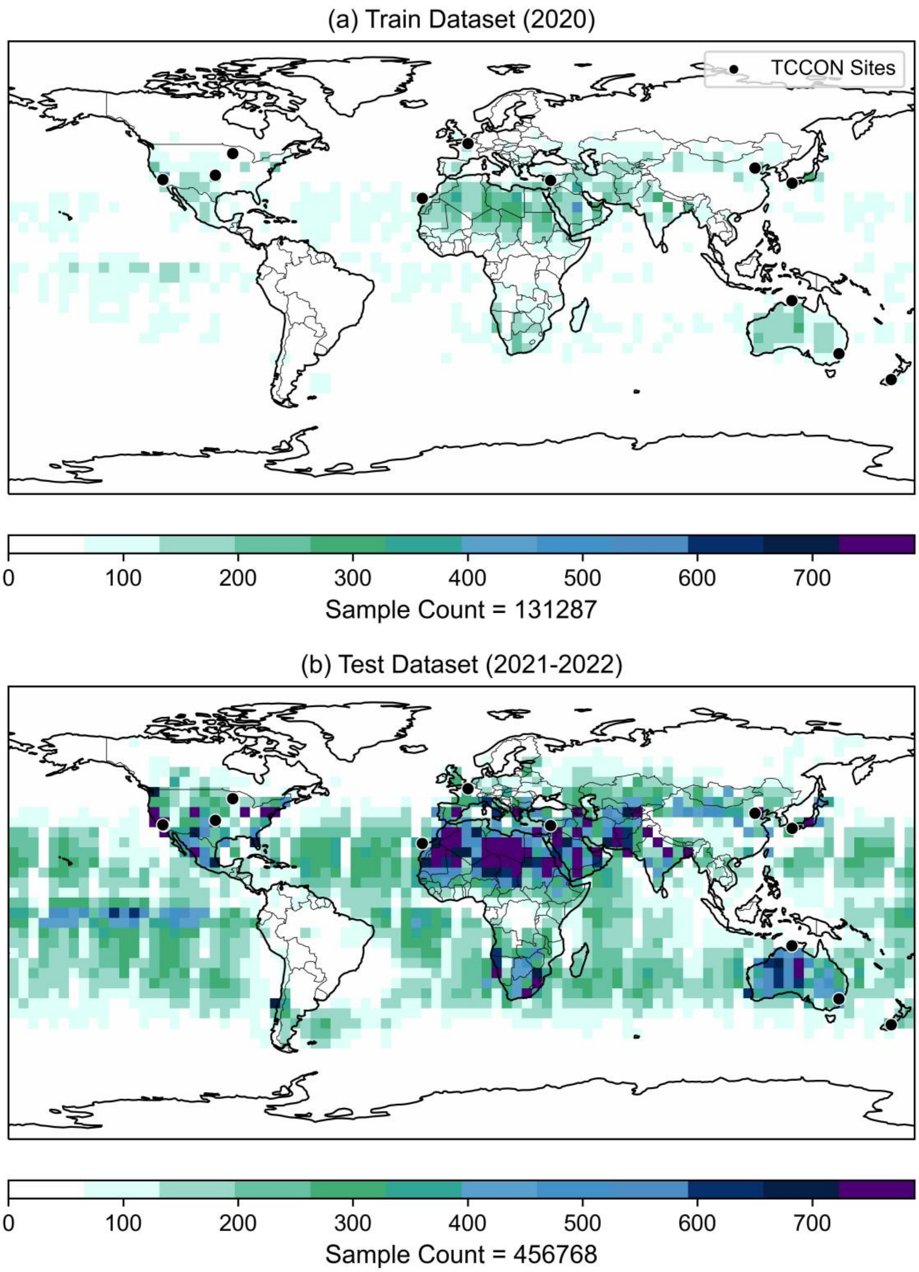


Figure 1. Spatial distribution of GOSAT-2 soundings used in this study. (a) Training data from 2020 used for ANN model development. (b) Independent prediction data from 2021–2022 used for model evaluation and global mapping. Colors indicate the number of valid soundings within each grid cell. Similar spatial coverage between the two periods reflects the

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

stable global sampling pattern of the GOSAT-2 observing system. Black markers indicate the locations of the TCCON stations used for validation.

3 Methods

3.1 ANN Retrieval Frameworks

Two ANN architectures were developed using different supervised training targets: one trained on FP XCH_4 (ANN-FP) and the other trained on proxy XCH_4 (ANN-Proxy). Both models take GOSAT-2 Level 1B observations as input but differ in network structure to reflect the information content emphasized by the corresponding retrieval target. The overall architectures of ANN-FP and ANN-Proxy are illustrated in Figure 2.

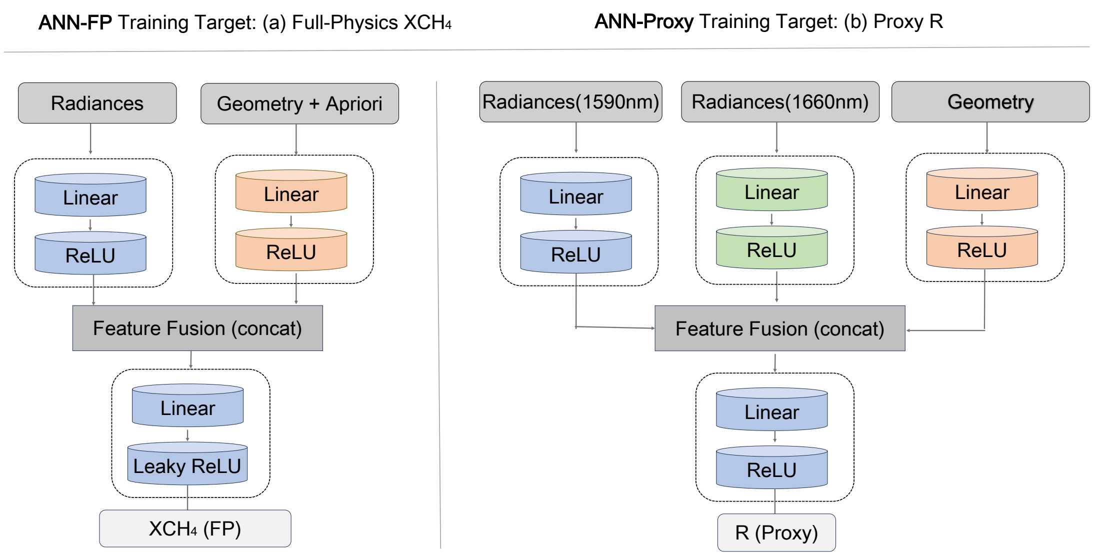


Figure 2. Model Architecture.

ANN-FP uses a two-branch fully connected network designed to approximate the coupled nonlinear mapping of the operational FP retrieval. The spectral branch maps high-dimensional SWIR radiances to gas absorption features using three hidden layers (1024, 512, and 512 units) with LeakyReLU activations. Concurrently, the auxiliary branch processes geometric and prior information variables through two hidden layers of 32 units each. The resulting feature representations from both branches are concatenated and passed through a fusion regression head, consisting of one hidden layer with 512 units, to predict XCH_4.

ANN-Proxy adopts a more compact three-branch architecture, where the physical structure is introduced through the supervised target definition and branch-wise spectral decomposition rather than explicitly encoding the proxy retrieval equation. Separate spectral branches independently process the CO_2 (1590 nm) and CH_4 (1660 nm) windows using three hidden layers each (512, 256, and 128 units) with ReLU activations, while a third geometric branch processes observing parameters via two hidden layers (32 and 16 units). These representations are

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

concatenated and mapped to the target ratio R through a 128-unit regression head. The final XCH4 estimate is then reconstructed by multiplying the predicted ratio by the corresponding model-based XCO_2 value.

Both models are trained by minimizing the mean squared error (MSE) between network predictions and the corresponding L2 training targets. Optimization is performed using stochastic gradient descent with adaptive learning-rate scheduling, and hyperparameters are selected based on validation performance.

For model development, the 2020 observations were partitioned into two temporally interleaved subsets (alternating by day), with one subset used for training and the other for in-sample validation. Observations from 2021–2022 were withheld entirely from the training and validation process and reserved exclusively for out-of-sample testing.

3.2 Environmental Perturbation Modes

To characterize the dominant spectral variability associated with Aerosol Optical Depth (AOD) and surface reflectance, we constructed empirical perturbation modes directly from the independent test data.

For a given environmental variable z(either AOD or albedo), observations were divided into the upper and lower 5th-percentile subsets according to the corresponding L2 retrieval parameter.

Let L^{high} and L^{low} denote spectra belonging to the high- and low-percentile groups, respectively.

Difference spectra were constructed as

$$\Delta L_i = L_i^{high} - L_i^{low} \quad (1)$$

Singular value decomposition (SVD) was then applied to the matrix of difference spectra, and the first right-singular vector was retained as the dominant environmental perturbation mode,

$$u = PC_1(\Delta L) \quad (2)$$

This procedure yields a data-driven estimate of the leading spectral variability associated with changes in the retrieved aerosol or surface-reflectance state within the observed GOSAT-2 sample distribution. Consequently, the resulting perturbation modes should be interpreted as empirical environmental variability modes embedded in the retrieval data set rather than physically prescribed radiative-transfer perturbations.

4 Results

4.1 Comparing With Operational retrievals

Figure 3 compares ANN retrievals with the operational GOSAT-2 L2 XCH_4 products under in-sample (2020) and out-of-sample (2021–2022) conditions. Under in-sample conditions (Figure 3a-b), both ANN-Proxy and ANN-FP agree closely with their respective L2 products, showing high correlations, near-unity regression slopes, and small biases. This consistency is also evident in the regional high-XCH_4 case (Figure 3c), where both models successfully reproduce the enhanced CH_4 pattern.

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

Under out-of-sample conditions, however, their performance diverges. ANN-FP (Figure 3d-e) shows reduced agreement with the operational FP product, including lower regression slope and larger standard deviation, indicating a partial loss of variability in newly acquired observations. This degradation is also reflected in the regional high-XCH_4 case, where ANN-FP no longer reproduces the enhanced methane feature clearly. This behavior is further supported by the monthly global distributions during the out-of-sample period. ANN-FP exhibits a progressively increasing negative bias and enhanced spatial heterogeneity in the second year, indicating degradation under extrapolation, whereas ANN-Proxy remains stable with consistent spatial patterns. In contrast, ANN-Proxy maintains stable performance across both in-sample and out-of-sample conditions, with only minor changes in the statistical metrics and continued robust representation of the regional high-XCH_4 pattern.

Overall, ANN-Proxy exhibits substantially more stable error characteristics and higher predictive stability than ANN-FP when evaluated against their respective operational Level 2 products. Its much smaller scatter (about 4-6 ppb, compared with about 9-11 ppb for ANN-FP) suggests that proxy XCH_4 defines a more stationary retrieval target under changing observational conditions.

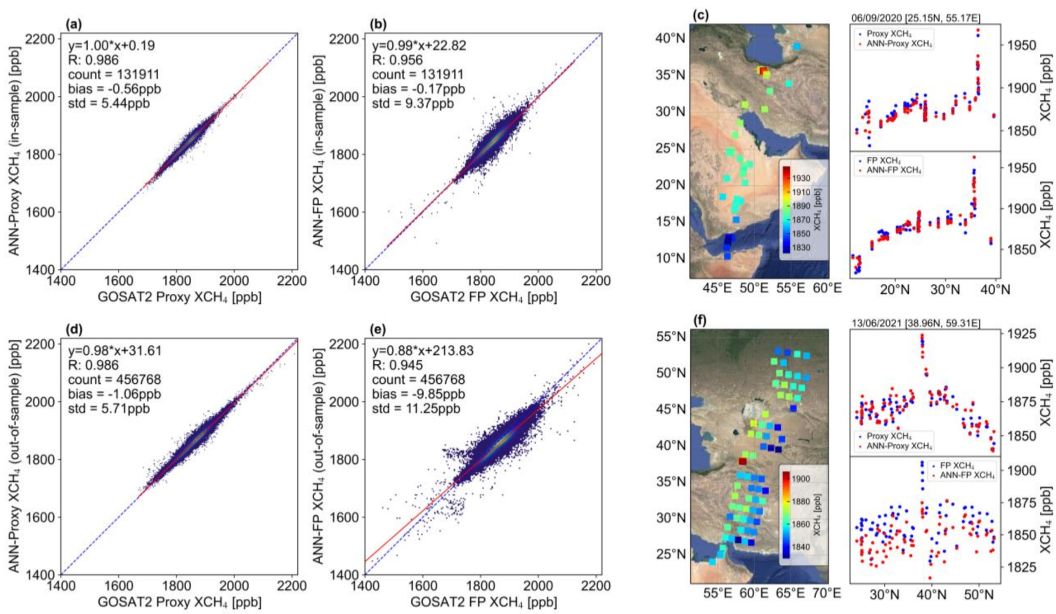


Figure 3. Evaluation of ANN retrievals against operational GOSAT-2 L2 XCH_4 products. (a–b) In-sample (within training period 2020) comparisons, including (a) ANN-Proxy vs. proxy XCH_4 and (b) ANN-FP vs. FP XCH_4. (c) Example of a high-emission point source on 6 September 2020 under in-sample conditions, showing consistent performance between ANN, proxy, and FP retrievals. (d–e) Out-of-sample (beyond training period 2021-2022) comparisons for ANN-FP vs. FP XCH_4. (f) Detection of a high-emission point source on 13 June 2021 under out-of-sample conditions. Solid lines

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

indicate linear regressions, and blue lines represent the 1:1 reference.

Figure 4 summarizes the monthly correlation coefficients and mean absolute errors between ANN predictions and the corresponding operational L2 products. Both ANN-FP and ANN-Proxy exhibit a recurring seasonal cycle, with the highest correlations generally occurring during boreal winter and the lowest values appearing during boreal summer. Similar seasonal behavior has been reported in previous GOSAT and TROPOMI studies (Bradley et al., 2025), where enhanced AOD, stronger atmospheric scattering, increased boundary-layer variability, and seasonal changes in surface reflectance during summer were shown to increase the complexity of SWIR methane retrievals (Butz et al., 2011; Schepers et al., 2012; Yoshida et al., 2011). Seasonal variations in surface albedo have also been identified as an important source of distribution shift for machine-learning-based methane retrievals, leading to reduced model transferability across seasons and observation periods (Bradley et al., 2025). The seasonal pattern observed here is therefore consistent with both traditional retrieval studies and recent deep-learning investigations of satellite XCH_4 retrieval performance.

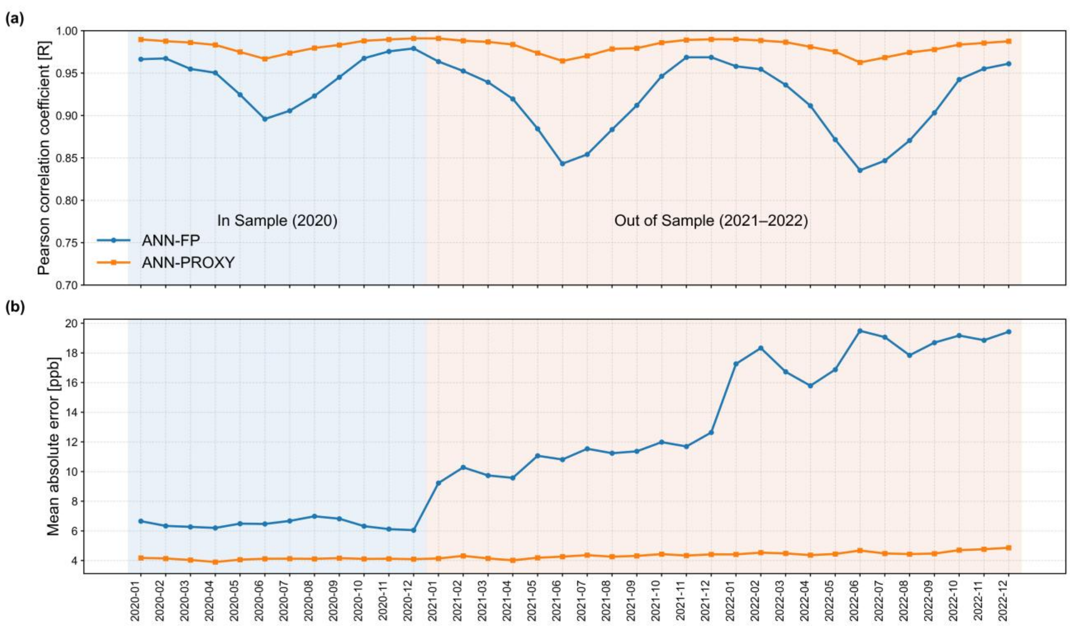


Figure 4. Temporal generalization performance of ANN-FP and ANN-Proxy. (a) Monthly Pearson correlation coefficients (R) between ANN predictions and operational L2 XCH_4 products. (b) Monthly mean absolute errors (MAE) in ppb. The light blue background (2020) indicates the in-sample period, while the light yellow background (2021–2022) represents the out-of-sample period used for independent testing.

In contrast, ANN-FP exhibits a progressive amplification of the seasonal degradation. The summer correlation minima decrease from approximately 0.90 during the training year to about 0.84 by 2022, while summer MAEs increase from roughly 6–7 ppb to nearly 20 ppb. More importantly, the degradation is not restricted to summer

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

conditions. Relative to 2020, elevated MAEs are observed during nearly all months of 2021–2022, including winter periods that are generally considered favorable for methane retrieval. This systematic drift suggests that the deterioration cannot be attributed solely to seasonal retrieval difficulty but instead reflects a gradual loss of temporal transferability as the observations become increasingly separated from the training period.

The contrasting behavior of the two models is consistent with fundamental differences between the FP and proxy retrieval frameworks. FP retrievals rely on a coupled inversion of methane, atmospheric scattering properties, and surface parameters, producing a target variable that is influenced by multiple interacting state variables and retrieval assumptions (Butz et al., 2011; Yoshida et al., 2011). Consequently, the mapping between Level-1 (L1) radiances and FP XCH_4 may evolve as atmospheric and surface conditions change. In contrast, proxy retrievals exploit the ratio of CH_4 and CO_2 absorption measurements, allowing a substantial fraction of common light-path errors to cancel (Parker et al., 2020; Schepers et al., 2012). The resulting proxy XCH_4 product therefore appears to define a comparatively more temporally stable target space, which can be learned and generalized more effectively by neural-network models. The distinct temporal behavior of ANN-Proxy and ANN-FP raises the question of whether their differences in extrapolation performance are also evident in the spatial distribution of retrieval errors. Figure 5 presents the annual mean residual fields for 2021 and 2022.

The residual patterns produced by ANN-Proxy show little change between the two years. Positive and negative residuals are distributed across most regions, and no persistent large-scale bias is apparent. Although weak hemispheric differences can be identified, with slightly more positive residuals over parts of the Northern Hemisphere and more negative residuals over portions of the Southern Hemisphere, the overall spatial structure remains largely unchanged from 2021 to 2022. This stability is consistent with the temporal error statistics and indicates that the ANN-Proxy retrieval retains a similar error structure under changing observational conditions.

The behavior of ANN-FP differs substantially. Negative residuals dominate much of the globe in both years and become more pronounced in 2022. The bias is particularly evident over continental regions, while oceanic areas generally exhibit smaller residual magnitudes. Nevertheless, the transition toward more negative residuals is observed across both land and ocean domains. The broad spatial extent of the bias suggests that the deterioration of ANN-FP is not confined to specific regions or surface types. Instead, it points to a systematic change in the mapping between L1 measurements and the FP retrieval target when the model is applied outside the conditions represented in the training data.

The spatial analysis therefore reinforces the temporal evaluation. ANN-Proxy preserves a balanced and geographically consistent residual distribution throughout the study period, whereas ANN-FP develops a widespread negative bias that strengthens over time. The coherence of this behavior across large spatial scales supports the view that proxy XCH_4 constitutes a more stationary target for statistical learning, leading to improved transferability and robustness under extrapolation conditions compared with FP XCH_4.

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

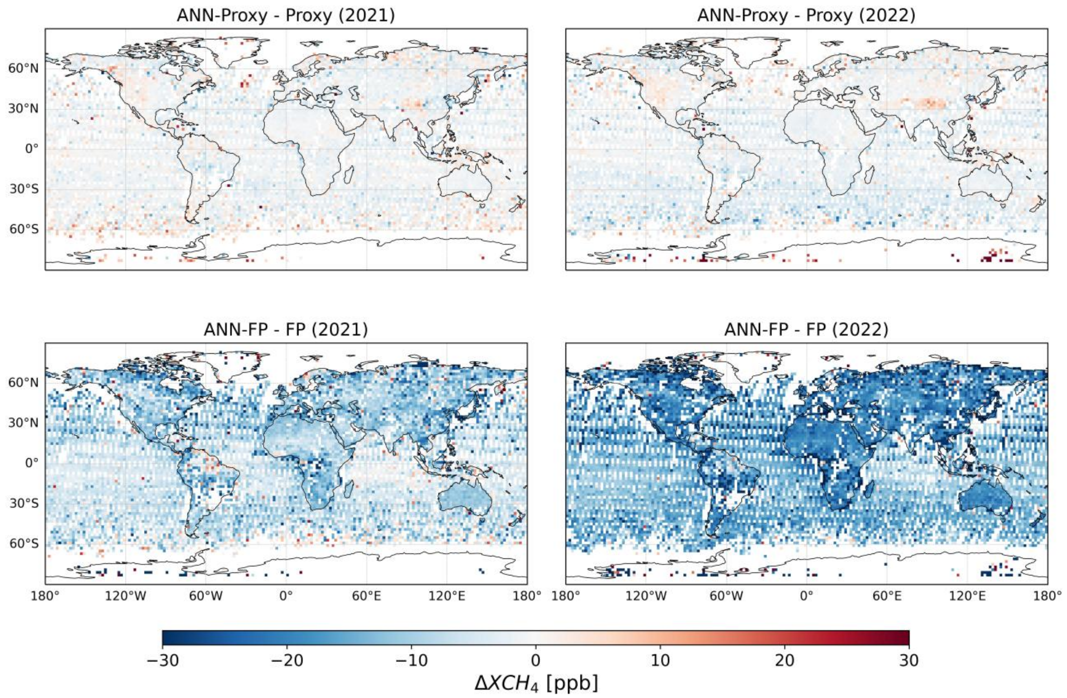


Figure 5. Annual mean residual fields for the out-of-sample period (2021–2022). (a)–(b) ANN-Proxy and (c)–(d) ANN-FP, relative to the operational GOSAT-2 L2 XCH_4 products.

Independent validation against collocated TCCON observations for 2020–2022 (Figure 6) provides an assessment of model performance relative to ground-based measurements rather than the operational L2 products. Consistent with the results obtained from the L2 comparisons, both ANN models reproduce a substantial fraction of the observed XCH_4 variability, but ANN-Proxy generally exhibits stronger agreement with TCCON across the validation network. At several sites, including Darwin, Izaña, Saga, and Xianghe, ANN-Proxy achieves higher correlations and smaller scatter than ANN-FP, whereas the differences between the two models are comparatively small at Caltech.

The TCCON evaluation further demonstrates that agreement with the operational retrieval product does not necessarily translate into superior performance against independent observations. Although ANN-FP closely reproduces the FP retrieval target, its performance degrades at several validation sites. In contrast, ANN-Proxy maintains more consistent behavior across diverse geographic and atmospheric conditions, indicating stronger robustness under independent evaluation.

Overall, the TCCON results are consistent with the temporal and spatial analyses presented above. The proxy-trained model not only exhibits superior extrapolation stability relative to the operational proxy product, but also achieves more reliable agreement with independent ground-based observations, supporting its suitability for rapid neural-network XCH_4 retrieval applications.

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

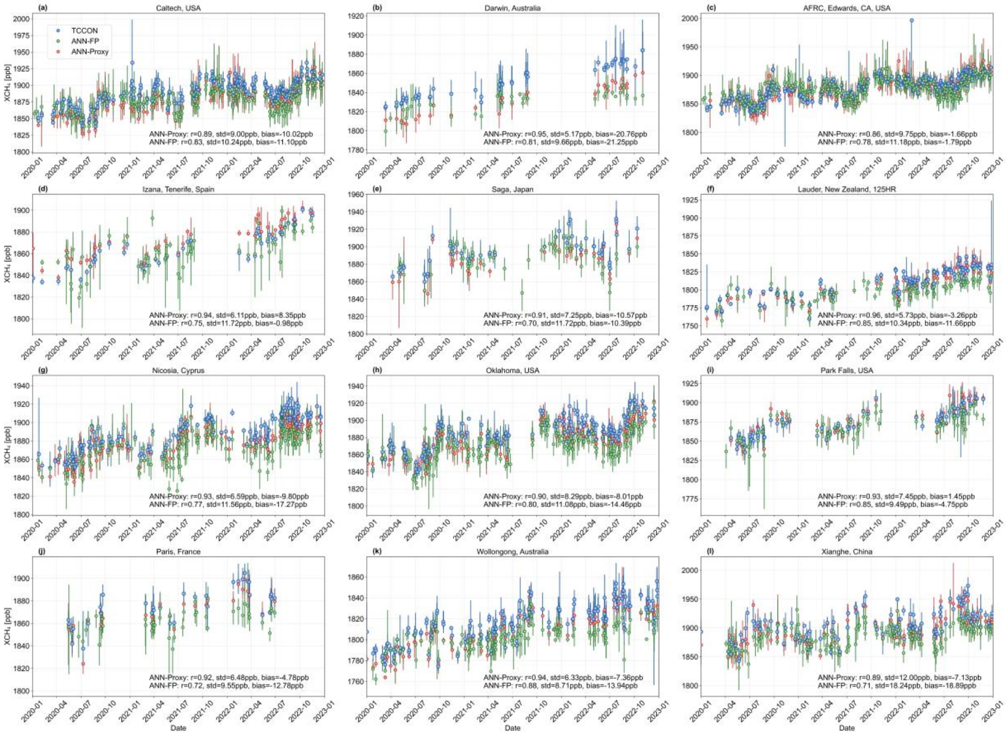


Figure 6. Site-by-site validation of ANN-FP and ANN-Proxy XCH_4 against ground-based observations from the TCCON (2020–2022).

4.2 Why Does the Proxy-Oriented Framework Exhibit Greater Robustness

The analyses above demonstrate that ANN-Proxy exhibits substantially stronger temporal transferability, spatial consistency, and agreement with independent TCCON observations than ANN-FP. Because the two retrieval frameworks differ not only in their supervised targets but also in the information pathways emphasized during learning, we investigate whether these performance differences are associated with distinct internal representations and sensitivity structures.

We first examine feature attribution using SHapley Additive exPlanations (SHAP). Although SHAP does not establish physical causality, it provides a useful diagnostic of how the networks utilize available information. As shown in Figure 7, ANN-FP distributes attribution across methane-sensitive radiances, viewing geometry, and prior-state variables, indicating a retrieval strategy that relies on coupled environmental and spectral information. In contrast, ANN-Proxy concentrates more than 97% of its attribution within the CH_4 and CO_2 absorption bands. This concentration is consistent with the proxy retrieval principle, in which the CH_4/CO_2 ratio suppresses a large fraction

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

of shared light-path variability before the learning process. The resulting representation therefore appears more strongly constrained by spectroscopic information and less dependent on auxiliary environmental variables.

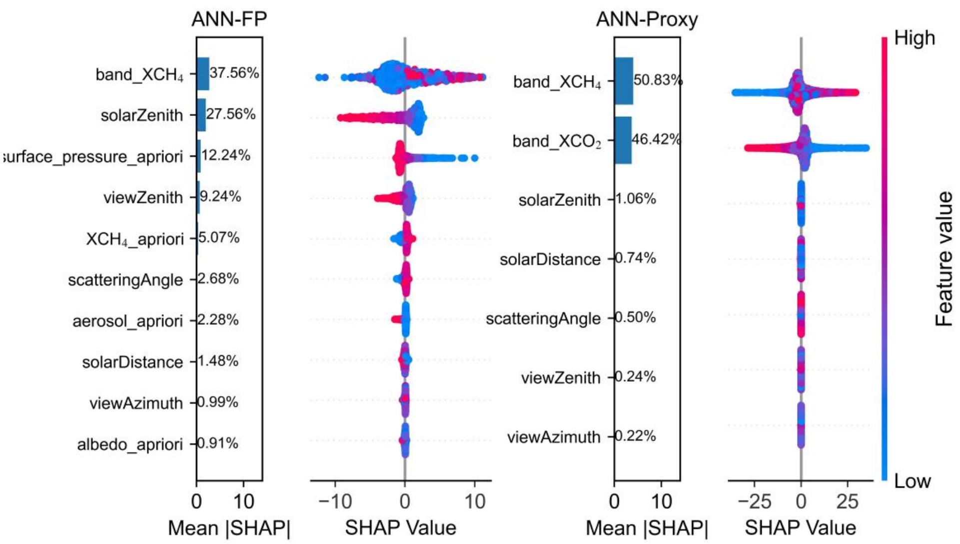


Figure 7. Internal attribution structures of ANN-FP and ANN-Proxy derived from SHAP.

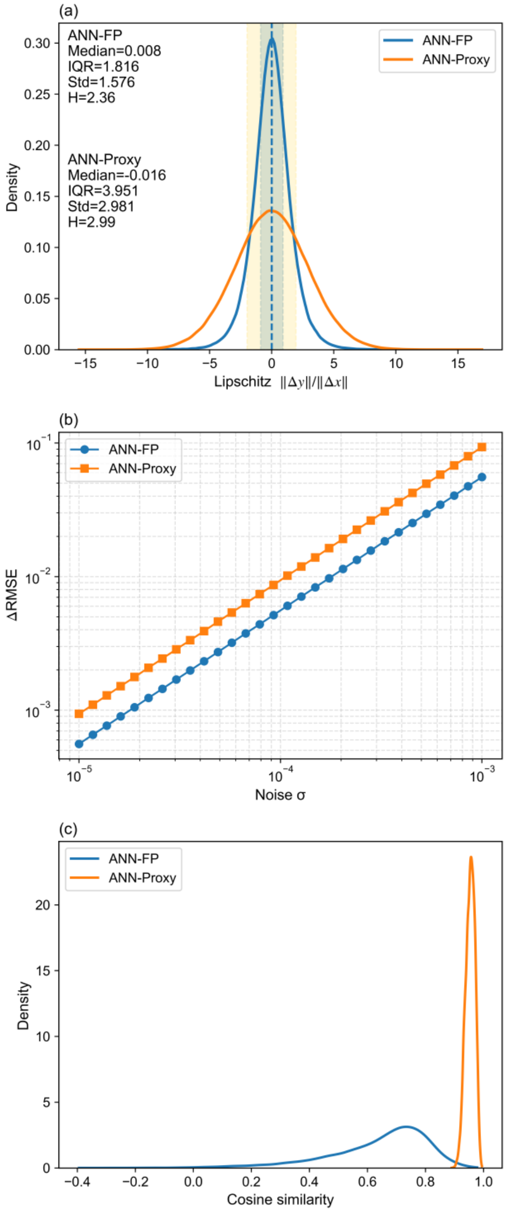
To further examine how these different representations translate into distinct sensitivity structures, we analyze the local geometry of the learned mappings. Local Lipschitz analysis (Figure 8a) indicates that ANN-Proxy operates within a broader sensitivity range than ANN-FP, consistent with its stronger reliance on information contained in the CH_4 and CO_2 absorption windows. In agreement with the SHAP results, ANN-Proxy gradients remain largely confined to the absorption subspace, whereas ANN-FP gradients are distributed across spectral, geometric, and prior-state variables.

This distinction is also reflected in the noise experiment (Figure 8b). When unstructured random perturbations are introduced into the spectral inputs, ANN-Proxy exhibits a larger response than ANN-FP, indicating greater sensitivity to perturbations applied directly within the spectral channels on which it primarily relies. However, this increased sensitivity does not imply a less stable mapping. Rather, the dominant sensitivity directions of ANN-Proxy remain remarkably consistent across samples, suggesting that information is concentrated within a smaller set of physically relevant pathways. In contrast, ANN-FP exhibits a weaker response to random spectral perturbations but a more distributed sensitivity structure involving multiple variable groups.

The gradient-direction cosine similarity provides the clearest evidence (Figure 8c). ANN-Proxy maintains an almost invariant principal sensitivity direction across the sample space (mean similarity = 0.96), whereas ANN-FP shows substantially greater directional rotation (mean similarity = 0.72). These results suggest that ANN-Proxy learns a

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

275 more structured and directionally consistent mapping, despite its stronger sensitivity to perturbations applied within
276 the absorption subspace.



277

278 **Figure 8. General mapping stability analysis. (a) Distribution of local Lipschitz constants. (b) Change in RMSE under**
279 **additive noise. (c) Pairwise cosine similarity of gradients across samples.**

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

However, geometric coherence alone does not guarantee robustness under physically meaningful retrieval variability. We therefore examine two dominant sources of uncertainty in SWIR methane retrievals, AOD and albedo, both of which alter atmospheric light-path characteristics and scene-dependent radiative conditions (Yu et al., 2024).

Under aerosol perturbations (Figure 9a–c), ANN-FP and ANN-Proxy exhibit markedly different behaviors. ANN-FP shows increasing retrieval variability with AOD and a broad gradient–perturbation alignment distribution, indicating substantial overlap between its dominant sensitivity directions and aerosol-induced spectral variability. By contrast, ANN-Proxy maintains nearly invariant retrieval statistics and remains concentrated near zero alignment, implying that aerosol-related perturbations project only weakly onto its dominant sensitivity directions.

Together with the SHAP results, these findings imply that ANN-FP retains stronger coupling between retrieval sensitivity and environmental state variables, whereas ANN-Proxy organizes sensitivity within a more restricted absorption-dominated subspace.

A similar pattern is observed for albedo perturbations (Figure 9d–f). ANN-FP exhibits systematic changes in both mean response and response variability across the albedo range, whereas ANN-Proxy remains considerably more stable. Consistent with the aerosol results, ANN-FP spans a broader gradient–perturbation alignment range and exhibits larger projection-dependent variability, while ANN-Proxy remains concentrated near zero alignment with comparatively uniform responses.

This behavior appears to arise from the proxy-oriented retrieval framework as a whole rather than from the retrieval target alone. The ratio-based physical constraint inherited from proxy retrieval theory, together with the concentration of learned sensitivity within a lower-dimensional absorption-dominated representation, yields a mapping that is less entangled with environmental perturbations and associated light-path uncertainties. These characteristics provide a plausible explanation for the superior temporal transferability and out-of-sample robustness of ANN-Proxy.Taken together, the aerosol- and albedo-conditioned analyses point toward a common mechanism: ANN-FP maintains stronger coupling between its dominant sensitivity directions and environmentally induced variability, whereas ANN-Proxy remains substantially more orthogonal to both perturbation manifolds.

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

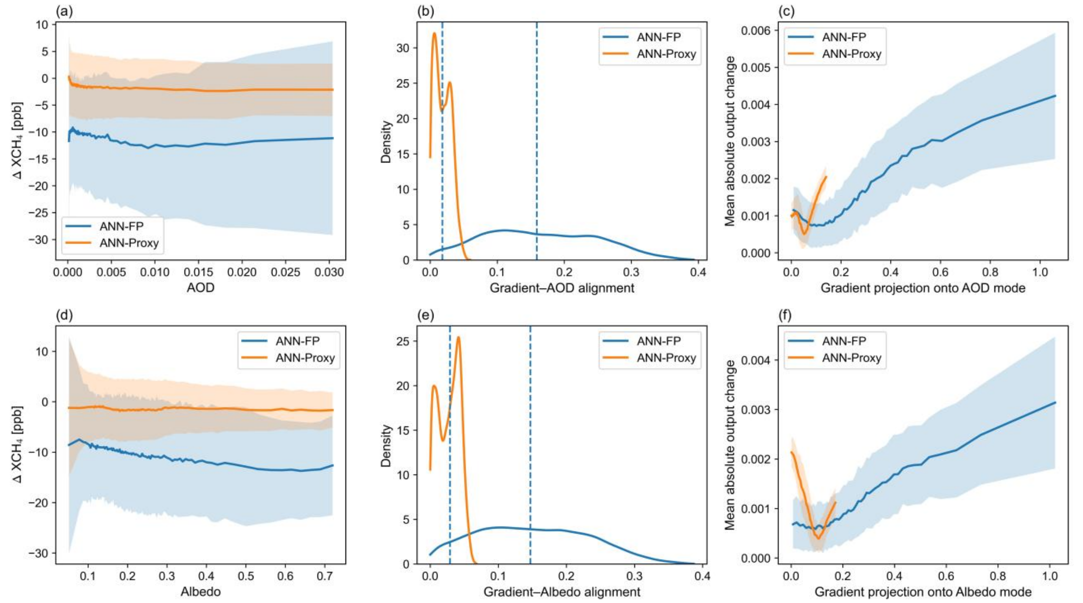


Figure 9. Sensitivity of ANN-FP and ANN-Proxy to environmental perturbations using out-of-sample observations. (a–c) aerosol-associated spectral mode analysis: (a) retrieval error versus AOD, (b) gradient–aerosol alignment, and (c) retrieval response versus aerosol-mode projection. (d–f) Surface-albedo-induced perturbation analysis: (d) retrieval error versus albedo, (e) gradient–albedo alignment, and (f) retrieval response versus albedo-mode projection.

5 Conclusions

This study examined how the choice of training target affects the structural stability and generalization of ANN-based methane retrievals. By comparing models trained on FP (ANN-FP) and proxy (ANN-Proxy) XCH_4 targets using GOSAT-2 observations, we show that models trained on proxy XCH_4 targets exhibit substantially stronger robustness under out-of-sample conditions. While both models reproduce their respective training targets with high fidelity, only ANN-Proxy maintains stable performance when applied to independent TCCON validation and unseen temporal domains.

Mechanistically, SHAP attribution reveals that ANN-FP relies on a high-dimensional compensatory coupling among spectral, geometric, and prior-related variables, whereas ANN-Proxy concentrates most of its attribution within the CH_4 and CO_2 absorption windows. Gradient, Lipschitz, and perturbation analyses further show that this proxy-oriented representation forms a more coherent local Jacobian structure, with sensitivity directions that are less entangled with aerosol-induced light-path perturbations and surface-reflectance variability. Together, these results suggest that the superior robustness of ANN-Proxy arises from a more constrained sensitivity geometry and weaker coupling to environmental variability.

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

From an operational perspective, ANN-Proxy is also substantially more compact and faster to evaluate. The proxy model size is 3417 kb, compared with 9206 kb for ANN-FP, and its inference time is approximately 0.01 ms per sample versus 0.03 ms per sample for ANN-FP on the same GPU platform. This combination of reduced model footprint, faster throughput, and improved out-of-sample stability makes ANN-Proxy a practical candidate for onboard methane processing in future carbon-monitoring satellites.

These findings suggest that physically structured target design can provide a practical pathway toward fast and robust machine-learning retrievals for future carbon-monitoring missions. For systems such as TanSat-2, which will generate substantially larger data volumes than current missions, proxy-oriented retrieval frameworks may help alleviate the computational burden associated with traditional physics-based inversions while maintaining robust performance under changing observation conditions. The superior robustness of the proxy-target retrieval arises from dimensional compression and the attenuation of light-path mode projection into the solution space. These findings highlight the critical value of physically structured target design for building fast, robust, and operationally deployable machine-learning retrievals. Future work may extend this target-design strategy beyond CH_4/CO_2 proxy retrieval. For example, oxygen could serve as an alternative proxy gas by combining the O_2 band near 1.27 μm with the CH_4 band near 1.6 μm, similar to the observing concept of MethaneSAT. Because O_2 provides information on optical path and dry air column, such targets may further improve the robustness and deployability of fast methane retrieval for next-generation carbon-monitoring satellites.

Data Availability Statement

The GOSAT-2 observations were obtained from the National Institute for Environmental Studies (NIES) via https://www.gosat-2.nies.go.jp/, as provided by Japan Aerospace Exploration Agency (JAXA), NIES, and the Ministry of the Environment, Japan (MOE). The TCCON data were obtained from the Total Carbon Column Observing Network via https://tccondata.org/, as described in D. Wunch et al. (2011). The retrieval codes and datasets used in this study are publicly available via Zenodo at https://doi.org/10.5281/zenodo.20095609.

Author contributions

Zhaonan Cai conceived and supervised the study. Rongjin Song developed the machine learning model, processed and analyzed the data, and prepared the manuscript. Yi Liu, Dongxu Yang, Lin Wu, Longfei Tian, Denghui Hu, and Guohua Liu contributed to the interpretation of the results and manuscript revision. All authors discussed the results and approved the final manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the National Key R&D Program of China (2022YFB3904802) and the Open Fund of the State Key Laboratory of Infrared Physics. The authors thank JAXA, NIES, and MOE for providing the GOSAT-2 data, and acknowledge TCCON for providing ground-based observations.

References

Aben, I., Hasekamp, O., & Hartmann, W. (2007). Uncertainties in the space-based measurements of CO2 columns due to scattering in the Earth's atmosphere. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *104*(3), 450–459. https://doi.org/10.1016/j.jqsrt.2006.09.013

Bradley, A. C., Dix, B., Mackenzie, F., Veefkind, J. P., & de Gouw, J. A. (2025). Deep transfer learning method for seasonal TROPOMI XCH_4 albedo correction. *Atmospheric Measurement Techniques*, *18*(7), 1675–1687. https://doi.org/10.5194/amt-18-1675-2025

Buchwitz, M., de Beek, R., Burrows, J. P., Bovensmann, H., Warneke, T., Notholt, J., et al. (2005). Atmospheric methane and carbon dioxide from SCIAMACHY satellite data: initial comparison with chemistry and transport models. *Atmospheric Chemistry and Physics*, *5*(4), 941–962. https://doi.org/10.5194/acp-5-941-2005

Butz, A., Guerlet, S., Hasekamp, O., Schepers, D., Galli, A., Aben, I., et al. (2011). Toward accurate CO2 and CH4 observations from GOSAT. *Geophysical Research Letters*, *38*(14). https://doi.org/10.1029/2011GL047888

David, L., Bréon, F.-M., & Chevallier, F. (2021). XCO2 estimates from the OCO-2 measurements using a neural network approach. *Atmospheric Measurement Techniques*, *14*(1), 117–132. Retrieved from https://amt.copernicus.org/articles/14/117/2021/amt-14-117-2021.html

Deutscher, N. M., Griffith, D. W. T., Bryant, G. W., Wennberg, P. O., Toon, G. C., Washenfelder, R. A., et al. (2010). Total column CO_2 measurements at Darwin, Australia – site description and calibration against in situ aircraft profiles. *Atmospheric Measurement Techniques*, *3*(4), 947–958. https://doi.org/10.5194/amt-3-947-2010

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

Feng, L., Palmer, P. I., Bösch, H., Parker, R. J., Webb, A. J., Correia, C. S. C., et al. (2017). Consistent regional fluxes of CH_4 and CO_2 inferred from GOSAT proxy $XCH_4 : XCO_2$ retrievals, 2010–2014. *Atmospheric Chemistry and Physics*, *17*(7), 4781–4797. https://doi.org/10.5194/acp-17-4781-2017

Hu, H., Landgraf, J., Detmers, R., Borsdorff, T., Aan de Brugh, J., Aben, I., et al. (2018). Toward Global Mapping of Methane With TROPOMI: First Results and Intersatellite Comparison to GOSAT. *Geophysical Research Letters*, *45*(8), 3682–3689. https://doi.org/10.1002/2018GL077259

Imasu, R., Matsunaga, T., Nakajima, M., Yoshida, Y., Shiomi, K., Morino, I., et al. (2023). Greenhouse gases Observing SATellite 2 (GOSAT-2): mission overview. *Progress in Earth and Planetary Science*, *10*(1), 33. https://doi.org/10.1186/s40645-023-00562-2

Jacob, D. J., Turner, A. J., Maasakkers, J. D., Sheng, J., Sun, K., Liu, X., et al. (2016). Satellite observations of atmospheric methane and their value for quantifying methane emissions. *Atmospheric Chemistry and Physics*, *16*(22), 14371–14396. https://doi.org/10.5194/acp-16-14371-2016

Jacot, A., Gabriel, F., & Hongler, C. (2018). Neural Tangent Kernel: Convergence and Generalization in Neural Networks. In *Advances in Neural Information Processing Systems* (Vol. 31). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper/2018/hash/5a4be1fa34e62bb8a6ec6b91d2462f5a-Abstract.html

Kuze, A., Suto, H., Nakajima, M., & Hamazaki, T. (2009). Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring. *Applied Optics*, *48*(35), 6716–6733. https://doi.org/10.1364/AO.48.006716

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Liang, A., Gong, W., Han, G., & Xiang, C. (2017). Comparison of Satellite-Observed XCO2 from GOSAT, OCO-2, and Ground-Based TCCON. *Remote Sensing*, *9*(10), 1033. https://doi.org/10.3390/rs9101033

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

Ma, X., Zhao, H., Jia, G., Xu, J., & Zhao, Y. (2026). DeepOE: A physics-informed neural network for atmospheric temperature profile retrieval. *Atmospheric and Oceanic Science Letters*, 100784. https://doi.org/10.1016/j.aosl.2026.100784

Monteil, G., Houweling, S., Butz, A., Guerlet, S., Schepers, D., Hasekamp, O., et al. (2013). Comparison of CH4 inversions based on 15 months of GOSAT and SCIAMACHY observations. *Journal of Geophysical Research: Atmospheres*, *118*(20), 11,807-11,823. https://doi.org/10.1002/2013JD019760

Morino, I., Uchino, O., Inoue, M., Yoshida, Y., Yokota, T., Wennberg, P. O., et al. (2011). Preliminary validation of column-averaged volume mixing ratios of carbon dioxide and methane retrieved from GOSAT short-wavelength infrared spectra. *Atmospheric Measurement Techniques*, *4*(6), 1061–1076. https://doi.org/10.5194/amt-4-1061-2011

Neyshabur, B., Bhojanapalli, S., Mcallester, D., & Srebro, N. (2017). Exploring Generalization in Deep Learning. In *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper/2017/hash/10ce03a1ed01077e3e289f3e53c72813-Abstract.html

O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., et al. (2012). The ACOS CO_2 retrieval algorithm – Part 1: Description and validation against synthetic observations. *Atmospheric Measurement Techniques*, *5*(1), 99–121. https://doi.org/10.5194/amt-5-99-2012

Pandey, S., Gautam, R., Houweling, S., van der Gon, H. D., Sadavarte, P., Borsdorff, T., et al. (2019). Satellite observations reveal extreme methane leakage from a natural gas well blowout. *Proceedings of the National Academy of Sciences*, *116*(52), 26376–26381. https://doi.org/10.1073/pnas.1908712116

Parker, R. J., Webb, A., Boesch, H., Somkuti, P., Barrio Guillo, R., Di Noia, A., et al. (2020). A decade of GOSAT Proxy satellite CH_4 observations. *Earth System Science Data*, *12*(4), 3383–3412. https://doi.org/10.5194/essd-12-3383-2020

Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., et al. (2019). On the Spectral Bias of Neural Networks. In *Proceedings of the 36th International Conference on Machine Learning* (pp. 5301–5310). PMLR. Retrieved from https://proceedings.mlr.press/v97/rahaman19a.html

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

Reuter, M., Hilker, M., Noël, S., Di Noia, A., Weimer, M., Schneising, O., et al. (2025). Retrieving the atmospheric concentrations of carbon dioxide and methane from the European Copernicus CO2M satellite mission using artificial neural networks. *Atmospheric Measurement Techniques*, *18*(1), 241–264. https://doi.org/10.5194/amt-18-241-2025

Schepers, D., Guerlet, S., Butz, A., Landgraf, J., Frankenberg, C., Hasekamp, O., et al. (2012). Methane retrievals from Greenhouse Gases Observing Satellite (GOSAT) shortwave infrared measurements: Performance comparison of proxy and physics retrieval algorithms. *Journal of Geophysical Research: Atmospheres*, *117*(D10). https://doi.org/10.1029/2012JD017549

Schneising, O., Burrows, J. P., Dickerson, R. R., Buchwitz, M., Reuter, M., & Bovensmann, H. (2014). Remote sensing of fugitive methane emissions from oil and gas production in North American tight geologic formations. *Earth's Future*, *2*(10), 548–558. https://doi.org/10.1002/2014EF000265

Suto, H., Kataoka, F., Kikuchi, N., Knuteson, R. O., Butz, A., Haun, M., et al. (2021). Thermal and near-infrared sensor for carbon observation Fourier transform spectrometer-2 (TANSO-FTS-2) on the Greenhouse gases Observing SATellite-2 (GOSAT-2) during its first year in orbit. *Atmospheric Measurement Techniques*, *14*(3), 2013–2039. https://doi.org/10.5194/amt-14-2013-2021

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., et al. (2011). The Total Carbon Column Observing Network. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1943), 2087–2112. https://doi.org/10.1098/rsta.2010.0240

Xie, F., Ren, T., Zhao, C., Wen, Y., Gu, Y., Zhou, M., et al. (2023). Fast retrieval of XCO2 over East Asia based on the OCO-2 spectral measurements. *Atmos. Meas. Tech.*, *17*, 3949–3967. https://doi.org/10.5194/amt-17-3949-2024

Yoshida, Y., Ota, Y., Eguchi, N., Kikuchi, N., Nobuta, K., Tran, H., et al. (2011). Retrieval algorithm for CO_2 and CH_4 column abundances from short-wavelength infrared spectral observations by the Greenhouse gases observing satellite. *Atmospheric Measurement Techniques*, *4*(4), 717–734. https://doi.org/10.5194/amt-4-717-2011

https://doi.org/10.5194/egusphere-2026-3549
Preprint. Discussion started: 31 August 2026

Yu, Q., Jervis, D., & Huang, Y. (2024). Accounting for the effect of aerosols in GHGSat methane retrieval. *Atmospheric Measurement Techniques*, *17*(11), 3347–3366. https://doi.org/10.5194/amt-17-3347-2024