

# A hybrid optimal estimation and machine learning approach to predict atmospheric composition

Frank Werner<sup>1</sup>, Kevin W. Bowman<sup>1</sup>, Seungwon Lee<sup>1</sup>, Joshua L. Laughner<sup>1</sup>, Vivienne H. Payne<sup>1</sup>, and James L. McDuffie<sup>1</sup>

<sup>1</sup>Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

**Correspondence:** Frank Werner (frank.werner@jpl.nasa.gov)

**Abstract.** We present a HYbrid REtrieval Framework (HYREF) that predicts subcolumn carbon monoxide (CO) concentrations from Cross-track Infrared Sounder (CrIS) observations, trained to replicate the TROpospheric Ozone and its Precursors from Earth System Sounding (TROPESS) retrievals based on optimal estimation (OE). Unlike the OE algorithm, which produces retrievals for only a small fraction of available CrIS observations due to computationally expensive but physically accurate radiative transfer, the addition of machine learning (ML) techniques enables full coverage by providing high-resolution predictions for every valid CrIS sample. Importantly, in addition to CO concentrations, TROPESS–HYREF also predicts key retrieval diagnostics, namely column averaging kernels, degrees of freedom, and retrieval errors, that are essential for meaningful comparison with other observations, models, and ingestion into data assimilation. The framework is designed to emulate and extend the OE retrieval, rather than replace it, by providing full spatial coverage and enhanced resolution consistent with the underlying physical solution.

The new framework achieves excellent performance with correlation coefficients  $r > 0.99$  and a bias  $< 0.1\%$  when benchmarked against an independent test set, and reproduces fine-scale spatial patterns in CO fields observed during a major wildfire over North America. A scale analysis reveals substantial variability in CO concentrations below the nominal  $0.80^\circ$  resolution of the TROPESS OE retrieval, which TROPESS–HYREF successfully resolves. Inference is computationally efficient, with daily global predictions completed in minutes on a single compute node. By filling observational gaps while maintaining consistency with the OE retrieval, this fusion of OE-derived physical information and ML-driven efficiency provides a practical pathway to high-resolution atmospheric CO monitoring with robust diagnostics.

*Copyright statement.* ©2025. California Institute of Technology.

## 1 Introduction

Carbon monoxide (CO) is a chemically reactive trace gas and key atmospheric pollutant, produced primarily through incomplete combustion of biomass and fossil fuels (Jacob, 1999), as well as through secondary production from the oxidation of methane (CH<sub>4</sub>) and non-methane hydrocarbons (e.g., Holloway et al., 2000). It plays a central role in atmospheric chemistry

by serving as a major sink for hydroxyl radicals (OH, Lelieveld et al., 2016), thereby influencing the oxidative capacity of the atmosphere and the lifetime of CH<sub>4</sub> (e.g., Gaubert et al., 2017). Due to its intermediate lifetime (weeks to months), CO serves  
25 as a valuable tracer for long-range pollution transport and chemical processing in the troposphere (e.g., Clerbaux et al., 2002; Edwards et al., 2004). It also contributes indirectly to radiative forcing via the formation of tropospheric ozone (O<sub>3</sub>) and carbon dioxide (CO<sub>2</sub>), classifying it as a short-lived climate pollutant (Bowman and Henze, 2012; IPCC, 2023).

Satellite observations of CO, beginning with the Measurements of Air Pollution from Satellites (MAPS, Reichle Jr. et al., 1990) in the early 1980s and continuing with instruments such as Measurement of Pollution in the Troposphere (MOPITT)  
30 (Drummond et al., 2010), Atmospheric Infrared Sounder (AIRS) (Aumann et al., 2003), Tropospheric Emission Spectrometer (TES) (Beer et al., 2001), Infrared Atmospheric Sounding Interferometer (IASI) (Clerbaux et al., 2009), Cross-track Infrared Sounder (CrIS) (Han et al., 2013), TROPOspheric Monitoring Instrument (TROPOMI) (Veefkind et al., 2012), Greenhouse Gases Observing Satellite 2 (GOSAT-2) (Noël et al., 2022) and Geostationary Interferometric Infrared Sounder (GIIRS) (Zeng et al., 2023), have provided a long-term, global perspective on CO distributions, emission sources, and trends (e.g., Worden  
35 et al., 2013; Buchholz et al., 2021). These datasets support air quality monitoring, inverse modeling of emissions, and evaluation of chemistry-climate models (e.g., Field et al., 2015, 2016; Buchholz et al., 2018). While global CO concentrations have declined over the past two decades due to improved combustion efficiency and decreased biomass burning (e.g., Schultz et al., 2015; Zheng et al., 2019), recent regional fire trends (see, e.g., Luo et al. 2024), and evolving air quality policies continue to shape CO variability, underscoring the need for sustained satellite observations with well-characterized uncertainties (e.g.,  
40 Strode et al., 2016). Nevertheless, changes in climate and extreme events can lead to substantial biomass burning events for which CO is a critical tracer to infer emissions (Byrne et al., 2021, 2024; Neyra-Nazarrett et al., 2025).

The NASA TROPOspheric Ozone and its Precursors from Earth System Sounding (TROPESS) project generates consistent, long-term records of tropospheric ozone and related trace gases, including CO (Bowman, 2021; Worden et al., 2022). Building on the TES legacy, TROPESS applies a unified optimal estimation (OE, see, e.g., Rodgers 2000) algorithm across multiple  
45 satellite platforms, supported by a comprehensive ground data system (Bowman et al., 2006; Fu et al., 2016). Emphasis is placed on rigorous uncertainty analysis and intercomparisons with independent observations to ensure the accuracy needed for trend detection. Figure 1a shows the spatial distribution of operational TROPESS Level 2 (L2) CO retrievals over the western United States on 10 June 2023, based on CrIS measurements. A regional zoom (red box) reveals that, due to computational constraints, only  $\approx 1.5\%$  of the available CrIS soundings are processed, leaving substantial gaps in global CO monitoring.

50 Machine learning (ML) approaches, whose use in atmospheric science has expanded in recent years (e.g., Grivas and Chaloulakou, 2006; Saponaro et al., 2013; Werner et al., 2020; Schultz et al., 2021; Werner et al., 2023), offer a promising path forward. ML models can efficiently learn complex, nonlinear relationships and provide rapid inference across large datasets. However, limitations in explainability and uncertainty quantification continue to hinder their broader application in remote sensing (Tyrallis and Papacharalampous, 2024).

55 In contrast to conventional OE retrievals, which produce not only the retrieved quantities of interest but also key diagnostics such as  $\chi^2$  statistics, degrees of freedom (DoF), retrieval precision, error covariance, and column averaging kernels (AK), ML

methods lack direct analogues to these quantities. Yet such diagnostics are critical for model–observation comparisons, data assimilation, and quality control (Jones et al., 2003; Miyazaki et al., 2015; von Clarmann and Glatthor, 2019).

Here, we present a novel hybrid framework that combines the strengths of OE and ML to generate high–resolution estimates of CO column concentrations from CrIS radiances. Our approach leverages OE retrievals as both training targets and sources of physically meaningful prior information, while enabling ML–driven capabilities such as rapid upscaling and the emulation of retrieval diagnostics. This fusion fills observational gaps left by current processing limits and provides an interpretable, uncertainty–aware pathway for incorporating ML into operational remote sensing pipelines. Importantly, the ML component is designed to emulate the OE retrieval and its associated diagnostics, rather than to replace or surpass the underlying physical solution, thereby extending OE–derived information to full spatial coverage. The TROPES–HYREF framework is therefore intended as a hybrid OE–ML system, in which the ML model operates alongside OE, for example by filling gaps in retrieval coverage, and can be periodically retrained as new OE results become available.

## 2 Data

The CrIS instrument, onboard NOAA’s Joint Polar Satellite System–1 (JPSS–1, also known as NOAA-20), is a Fourier Transform Spectrometer that captures Earth views across 30 cross–track interferograms, providing a swath width of 2,200 km. Each interferogram contains a  $3 \times 3$  array of fields of view (FOVs), with each circular FOV having a diameter of 14 km at nadir. CrIS data are processed to provide calibrated Level 1B (L1B) radiances in three spectral bands: 660–1095  $\text{cm}^{-1}$  (longwave), 1210–1750  $\text{cm}^{-1}$  (midwave), and 2155–2550  $\text{cm}^{-1}$  (shortwave). The instrument unapodized spectral resolution is 0.625–2.5  $\text{cm}^{-1}$ . NASA’s version 2 L1B radiances are available from the Goddard Earth Sciences Data and Information Services Center (GES DISC) UW-Madison Space Science and Engineering Center: Hank Revercomb; UMBC Atmospheric Spectroscopy Laboratory: Larrabee Strow, 2018.

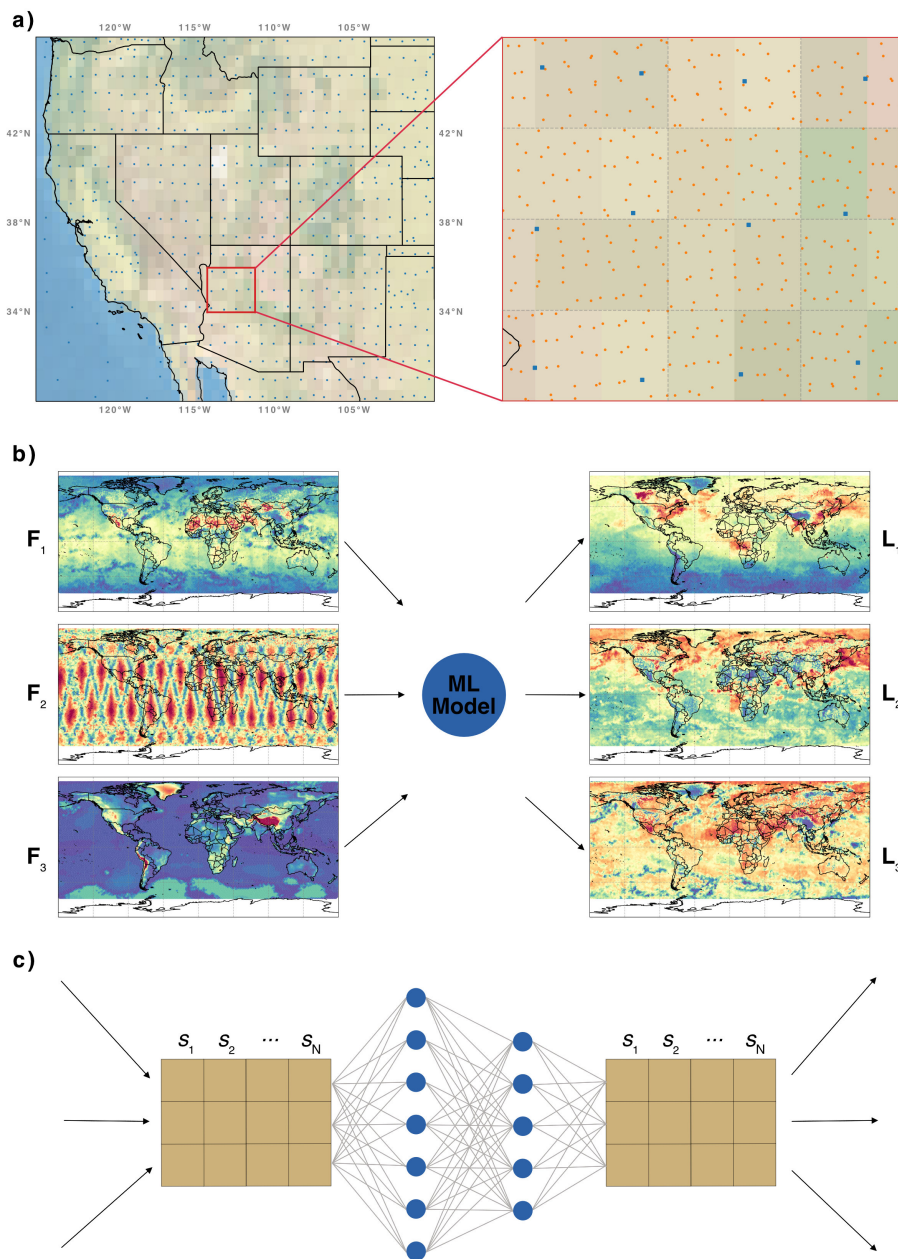
TROPES trace gas retrievals are provided on a reduced horizontal grid of  $0.8^\circ$  by the MUSES data processing system (Fu et al., 2016, 2018, 2019). These retrievals are based on the TES L2 processing algorithm (Bowman et al., 2006) and utilize an OE retrieval approach (Rodgers, 2000). TROPES retrievals of carbon monoxide (CO) are processed operationally, have undergone extensive verification (e.g., Worden et al., 2022; Luo et al., 2024), and are accessible via the GES DISC. In this study, single–FOV CrIS–MUSES retrievals from the TROPES forward stream were used (Bowman, 2021).

Data in this study is comprised of CrIS and TROPES data over 04/2023–01/2025.

## 3 ML model

### 3.1 Setup and training

We developed, trained, and evaluated a ML model to simultaneously predict a variety of TROPES CO variables, primarily using observed CrIS radiances and geolocation data as inputs. This setup is illustrated in the simplified diagram in Fig. 1b, where we drastically limit the input and output variables to aid visibility. In this example the model uses three features ( $\mathbf{F}_{1-3}$ )



**Figure 1.** (a) Geolocations of L2 CO retrievals (blue dots) and L1B CrIS radiances (orange dots) over the western United States on 10 June 2023. (b) Simplified sketch of the ML setup, where three features ( $F_{1-3}$ ; radiances at  $2,181.88 \text{ cm}^{-1}$ , sensor viewing angle, and surface altitude) are used as input for the ML model in order to predict three labels ( $L_{1-3}$ ; CO concentrations, retrieval error, and an individual column averaging kernel). (c) Simplified sketch of the ML model. The variables  $F_{1-3}$  are converted to a two-dimensional input matrix which connects to neurons in two hidden layers, and map to a two-dimensional output matrix, which provides  $L_{1-3}$ .

as input: CrIS radiances at  $2,181.88 \text{ cm}^{-1}$ , the sensor viewing angle, and the surface altitude, respectively. These features are matrices, where each element  $f_{1-3}^s$  corresponds to one of the  $N$  samples, indexed as  $s = 1, \dots, N$ . The ML model maps these features to a set of output labels ( $\mathbf{L}_{1-3}$ ), which in this simplified example are the CO total column concentrations, the total column retrieval error, and the column AK at  $\approx 511 \text{ hPa}$ . Like the features, these labels are matrices that contain elements  $l_{1-3}^s$  for each individual sample. Again,  $s = 1, \dots, N$  denotes the individual sample (i.e., CrIS column).

The ML model developed in this study is a feedforward artificial neural network (ANN), which maps the input to the output through several hidden layers, each consisting of a large number of interconnected neurons. A simplified schematic of an example ANN, with two hidden layers containing 7 and 5 neurons, respectively, is shown in Fig. 1c. This diagram also illustrates how the geolocated features are transformed into two-dimensional input and output matrices and how they connect to the individual neurons.

The exact model structure and hyperparameters (i.e., model settings) are determined through the procedures described in Werner et al. (2021, 2023). By applying  $k$ -fold cross-validation across a range of potential model setups, the ideal hyperparameters were found to be two hidden layers with 1506 neurons per layer, "Rectified Linear Unit" activation functions after each hidden layer, an L2 weight decay parameter of  $5^{-34}$ , and the "Adaptive Moment Estimation" optimizer with a learning rate of  $1e^{-5}$ . The loss function minimized during training is the mean squared error. For each training iteration, batches of samples are passed through the model in a forward pass to compute predictions, followed by a backward pass in which model weights are updated via backpropagation. Each mini-batch contains 8,192 samples. Further details on these parameters and their impact are provided in Reed and Marks (1999); Goodfellow et al. (2016); Werner et al. (2021).

Model training was carried out using the "Keras" library for Python (version 2.10.0; Chollet et al., 2015), with "TensorFlow" (version 2.10.0) as the backend (Abadi et al., 2016). Of the available CrIS radiances and TROPES retrieval over 04/2023–01/2025, 98% of randomly selected samples were used as training data. After each training iteration, the model's performance was evaluated for an independent validation dataset comprised of 1% of the available data (approximately 185,000 samples). After several thousand iterations, the model weights corresponding to the best performance scores on the validation set were saved.

The specific features used for the CO model include radiances from all 2,224 spectral channels, the FOV index, the latitude and longitude of each sample, UTC time, a day/night flag, the sensor viewing angle, the day of the year, and the TROPES subcolumn a priori values. This yields an input matrix containing 2,235 variables. Note that the surface altitude was included for models predicting retrievals and diagnostics for other TROPES species. We tested reduced channel sets focused on CO-sensitive regions but found that using the full spectrum provided slightly improved performance, likely due to additional information on atmospheric state variables (e.g., temperature and humidity). The predicted labels of the CO model consist of the subcolumn concentrations, column AKs, and subcolumn retrieval errors, resulting in an output matrix containing 24 variables.

Prior to training, these inputs and outputs were filtered to remove invalid samples using a set of basic quality filters, including non-finite values, fill values, extreme outliers, failed retrieval quality flags, and target values outside the valid retrieval range.

In addition, extreme outliers in the label distributions were masked using percentile-based tail cutoffs, and both input features and output labels were standardized before training.

Model training was performed on a high-performance computing cluster and took  $\approx 10$  days to converge to a solution for the  $> 12,000,000$  model weights.

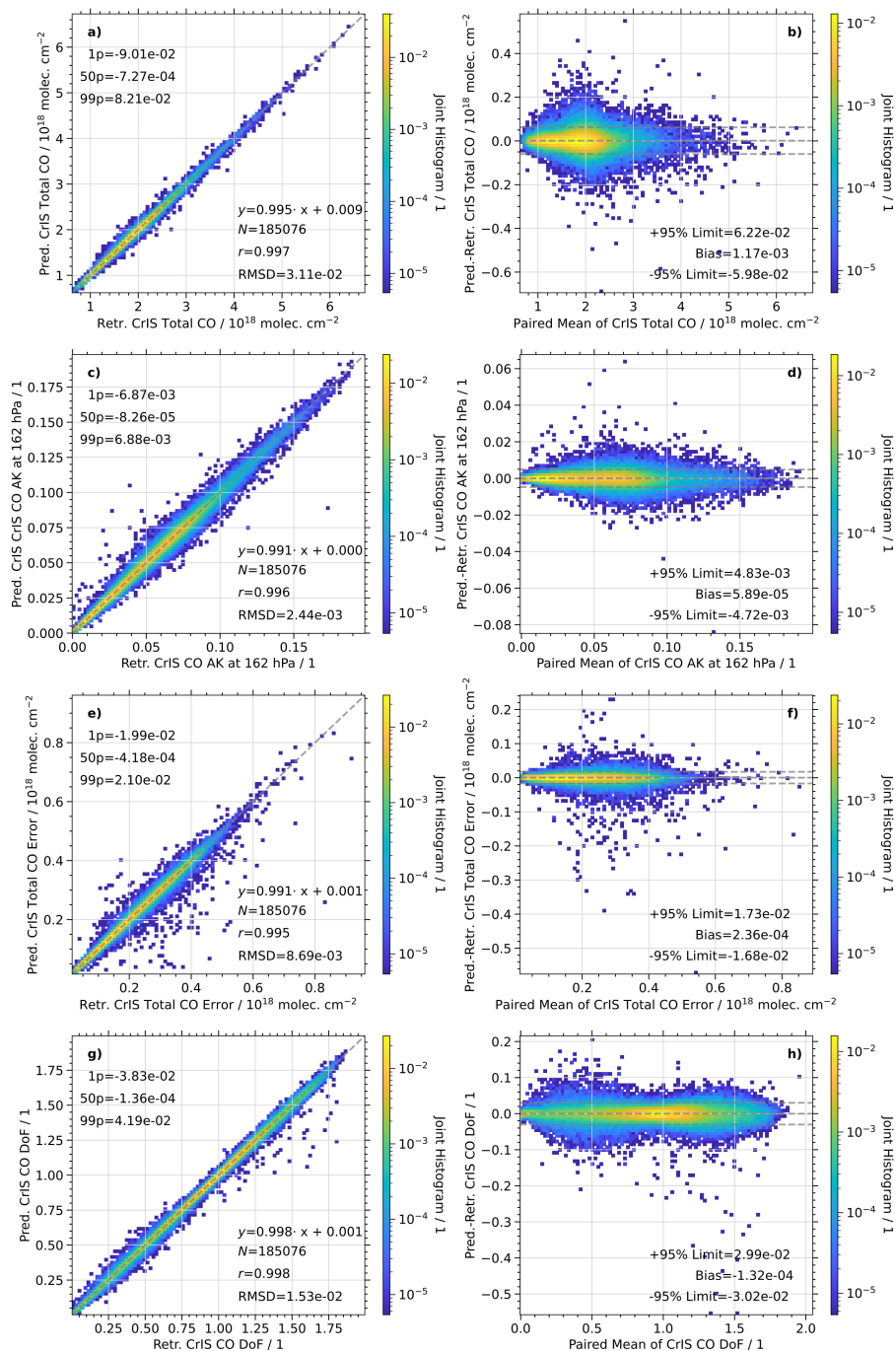
### 3.2 Evaluation

Model performance is evaluated using an independent test dataset, which consists of the remaining 1% of randomly sampled data that were not included in the training or validation process. Ideally, (i) the model should reliably predict CO concentration retrievals and OE diagnostics for these data points, even though the ML algorithm was not trained on them, and (ii) performance metrics should be similar to those derived from the training and validation datasets. It is important to note that the objective of the ML model is not to generalize beyond the statistical characteristics of the OE retrieval, but to emulate the OE solution and provide full spatial coverage for the same observing system. As such, the random split into training, validation, and test datasets ensures that the model is evaluated on samples that were not explicitly used during training, while still reflecting the same underlying distribution of atmospheric states and observational conditions. Given the strong spatial and temporal correlations inherent in satellite observations, this approach is appropriate for assessing the model's ability to reproduce OE retrievals across the full range of conditions encountered in the dataset. In operational use, the model is continuously retrained with newly available OE retrievals, ensuring consistency with evolving atmospheric variability.

Figure 2a presents a joint histogram of total column CO from the ML and OE algorithms for over 180,000 samples in the independent test dataset. Yellow colors represent regions with the highest density of data points, while blue colors correspond to areas with very few samples. The good agreement between the ML and OE results is evident, as most observations are narrowly clustered around the 1:1 line (indicated by the gray, dashed line). Five performance metrics are provided in the panel: Pearson's product-moment correlation coefficient ( $r$ ), the root-mean-square deviation (RMSD), the median deviation between the predicted and retrieved CO (50p, i.e., the bias), and the 1<sup>st</sup> and 99<sup>th</sup> percentiles of the deviation. Notably, for the total column concentrations, we find  $r > 0.99$ ,  $\text{RMSD} = 3.11 \times 10^{16}$  molecules  $\text{cm}^{-2}$ , a median difference of  $-7.27 \times 10^{14}$  molecules  $\text{cm}^{-2}$ , and maximum absolute differences for the majority of samples of  $< 1.00 \times 10^{17}$  molecules  $\text{cm}^{-2}$ .

Similar comparisons for tropospheric column concentrations, total AK at 162 hPa, and DoF are shown in Figs. 2c, e, and g. Again, the distributions closely follows the 1:1 line, with similarly high correlations ( $r > 0.99$ ). The lowest correlation occurs for the column AK at the lowest atmospheric level (not shown), where  $r = 0.98$ . These performance metrics are almost identical to those obtained for the validation dataset, where the comparison of predicted and retrieved total column CO concentrations yields  $r > 0.99$ ,  $\text{RMSD} = 3.14 \times 10^{16}$  molecules  $\text{cm}^{-2}$ , a median difference of  $6.57 \times 10^{14}$  molecules  $\text{cm}^{-2}$ , and maximum absolute differences for the majority of samples of  $< 1.00 \times 10^{17}$  molecules  $\text{cm}^{-2}$ .

Performance metrics are consistent across training, validation, and test datasets, with nearly identical correlation coefficients ( $|\Delta r| < 0.0006$ ), normalized RMSD values ( $|\Delta \text{RMSD}| < 0.21\%$ , apart from the last two column AK levels deep in the stratosphere where values of effectively 0), and biases ( $|\Delta \text{bias}| < 0.12\%$ ) for all predicted variables. This indicates that there is no evidence of overfitting and that the model exhibits stable behavior across the available datasets.



**Figure 2.** (a) Joint histogram and (b) Bland-Altman plot of predicted and retrieved total column CO concentrations from the test data set. (c)–(h) Similar to (a) and (b), but for the column averaging kernel (AK) at 162 hPa, total column retrieval error, and degrees of freedom (DoF).

We note that the use of a random split does not enforce strict independence between training, validation, and test datasets, as atmospheric states exhibit strong spatial and temporal correlations. In this study, the purpose of the split is therefore not to assess fully independent generalization, but to evaluate how well the model reproduces OE retrieval behavior across the distribution of atmospheric states and viewing geometries sampled by the instrument.

160 To ensure that the split remains representative, we verified that the distributions of key variables are statistically consistent across training, validation, and test datasets using Kolmogorov–Smirnov tests. Combined with the consistent performance metrics reported above, this indicates that the model generalizes well within the sampled observational distribution. This behavior is consistent with the intended hybrid OE—ML application, in which the model is designed to operate on the same observational distribution as OE.

165 To complement the regression analysis, we employ Bland–Altman plots to further assess the agreement between the OE and predicted results. This approach highlights systematic differences and potential heteroscedasticity that may not be apparent in standard correlation-based evaluations, especially for non-Gaussian or magnitude-dependent variability, and has been shown to provide a more robust framework for intercomparison of geophysical datasets (e.g., Knobelspiesse et al., 2019). Panels b, d, f, and h show the Bland–Altman distributions, with the paired mean of predicted and OE values on the x-axis and the difference  
170 between the two datasets on the y-axis, along with three horizontal dashed lines. The central line denotes the mean difference (bias), while the outer lines show the 95% limits of agreement (mean  $\pm 1.96$  standard deviations), indicating the interval that contains approximately 95% of the differences under the assumption of normally distributed residuals.

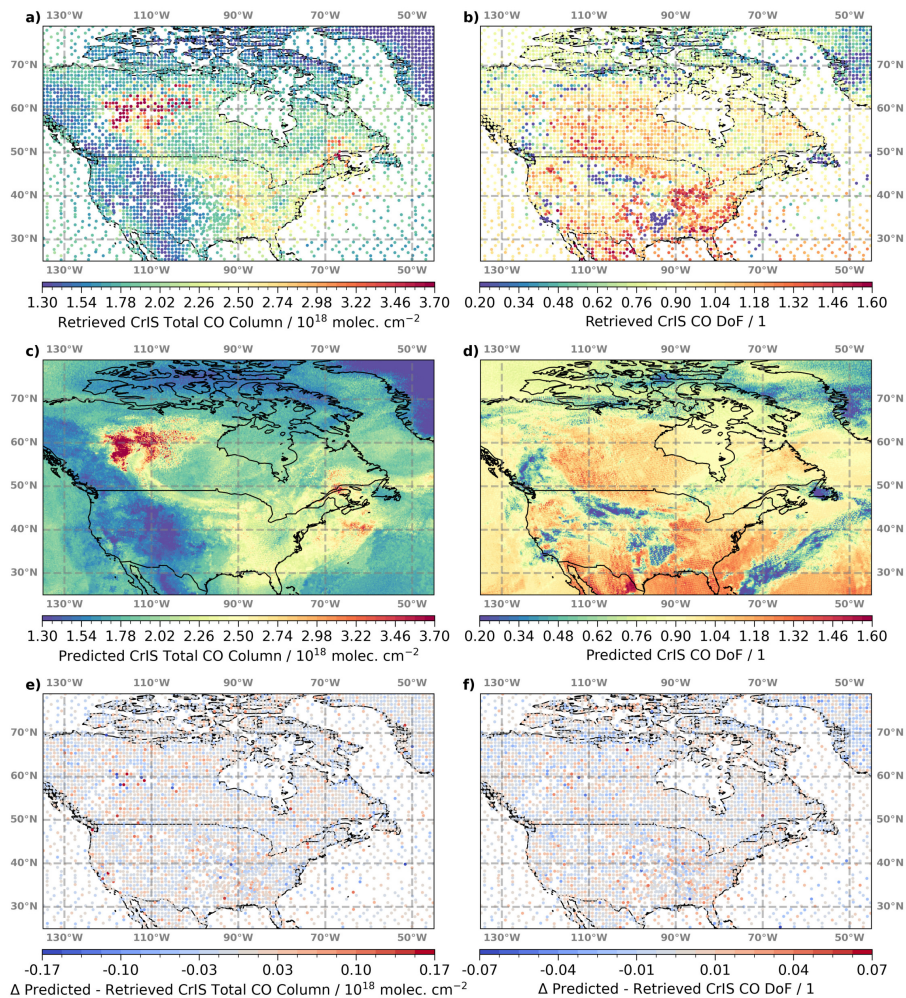
The distributions show no evidence of magnitude-dependent bias or systematic slope. The proportion of data points within the 95% limits of agreement (95%, 95%, 96%, and 95%, respectively) aligns closely with expectations for normally distributed  
175 residuals, and 79%, 80%, 86%, and 78% of points fall within  $\pm 1$  standard deviation. This indicates well-behaved error distributions with no pronounced evidence of heavy tails or heteroscedastic spread, suggesting that the model adequately captures variability across the full dynamic range. This is particularly relevant for machine learning retrievals, which can otherwise struggle in the presence of heteroscedastic relationships between observables and geophysical variables (e.g., Miller et al., 2020), and indicates that such effects are not evident in our ML predictions. Minor increases in spread near the peak of the  
180 distributions reflect regions of highest data density and do not indicate systematic bias or magnitude-dependent variability.

The close agreement between predicted and OE-derived diagnostics indicates that the ML model not only reproduces the retrieved state, but also captures the associated sensitivity and uncertainty characteristics of the OE solution. This consistency is critical for downstream applications, such as data assimilation and model evaluation, where the proper interpretation of retrievals depends on the availability of reliable AKs and error estimates.

## 185 4 Results

### 4.1 Example maps

Figure 3a presents a representative example scene of total column CO from the TROPES OE retrieval on 10 June 2023. A large area of enhanced CO concentrations ( $> 3 \times 10^{18}$  molecules  $\text{cm}^{-2}$ ) is evident over Western Canada, associated with the



**Figure 3.** (a) Example scene of OE retrievals of total column CO over North America on 10 June 2023. (b) Similar to (a), but showing the associated OE degrees of freedom (DoF). (c)–(d) Similar to (a)–(b), but for the ML predictions. (e)–(f) Differences between colocated ML predictions and OE retrievals, and their respective error estimates.

unprecedented wildfire season that year (Jain et al., 2024). These fires produced large smoke plumes that affected portions of Canada and the United States for several weeks, before spreading across the Northern Hemisphere. Notably, enhanced CO concentrations ( $> 2.5 \times 10^{18}$  molecules  $\text{cm}^{-2}$ ) are also recorded over Eastern Canada, the entire Eastern United States, and parts of the Atlantic Ocean.

The associated OE DoF are shown in Fig. 3b. Areas of moderate to high CO concentrations generally coincide with regions of elevated DoF. Smaller DoF  $< 0.6$  are observed over Greenland, the Atlantic Ocean, and over isolated regions over the continental United States. These reduced DoF are indicative of lower retrieval sensitivity and are likely due to the interference of clouds or poor thermal contrast.

Figures 3c–d show the ML predictions for total CO concentrations and DoF, respectively. These results are derived for each CrIS L1B sample. The increased spatial resolution is particularly noticeable over the oceans, but even over land the ML results capture much finer spatial features, while faithfully reproducing the CO enhancements and DoF from the OE retrieval. Divergence maps in Figs. 3e–f illustrate the differences between predicted and retrieved results. The median differences are < 0.1% for both variables, and for the majority of samples (i.e., within the 5<sup>th</sup> and 95<sup>th</sup> percentiles), ML predictions are within  $\pm 2.40\%$  for total CO concentrations and within  $\pm 4.12\%$  for DoF. Overall, the difference between ML and OE total column CO concentrations exceeds the retrieval error for only 14 of the 5,308 samples in the scene (0.26%). Similarly, excellent agreement is observed for the retrieval errors (not shown), with a majority of ML predictions within  $\pm 6.11\%$  and a median difference of 0.04%.

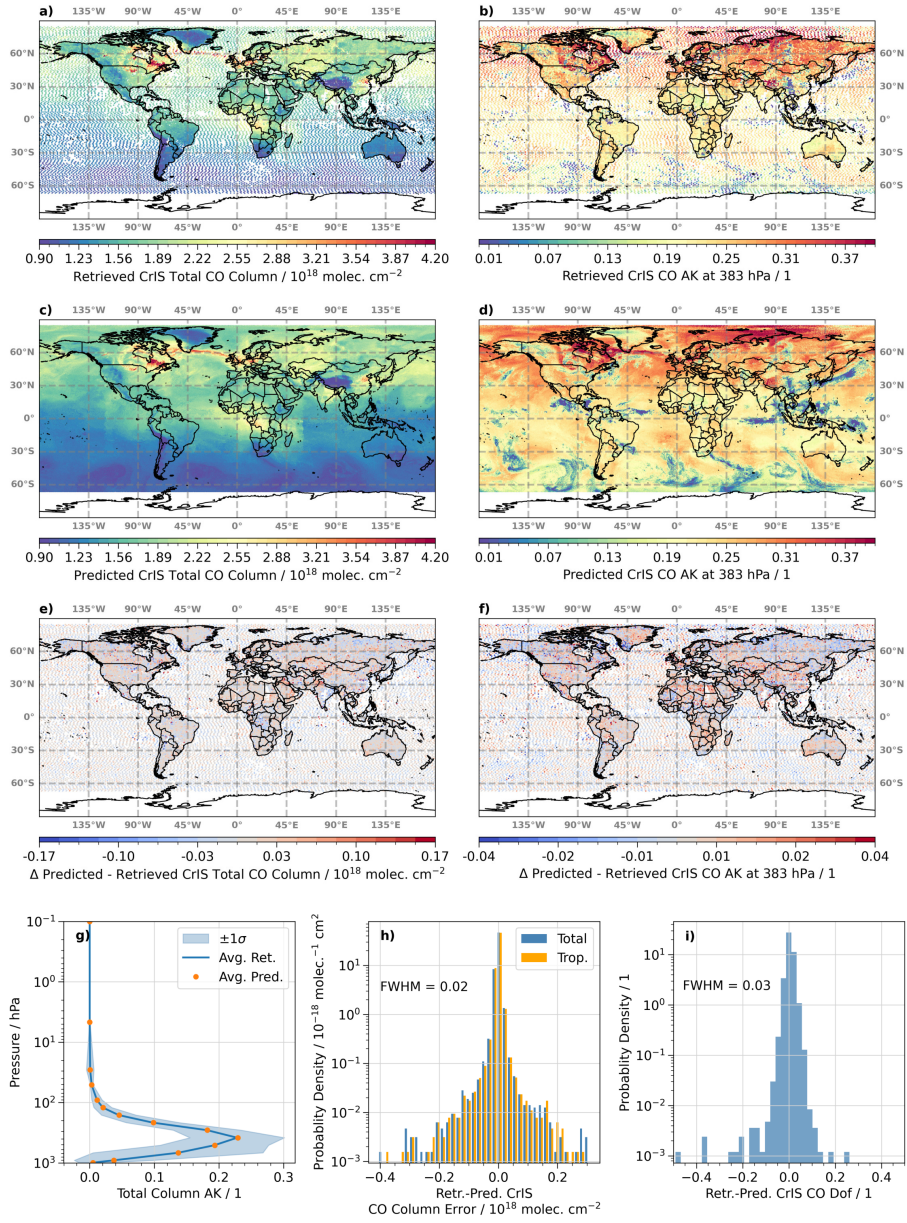
The close agreement between the ML predictions and OE retrieval shown in Fig. 3 primarily reflects interpolation within the sampled observational distribution, as this day is part of the training period, rather than fully independent spatiotemporal generalization. To further assess the performance of the ML model beyond the training period, Fig. 4 shows a global scene for 8 June 2025, which lies outside the training range (04/2023–01/2025). This experiment represents a limited temporal extrapolation test rather than a comprehensive assessment of long-term model stability. The results demonstrate that the model retains strong predictive skill under these conditions, indicating that it can generalize to unseen temporal states to a certain extent. However, we emphasize that such standalone predictive capability is not the primary objective of the framework. The model is designed to operate as part of a hybrid OE–ML system, benefiting from periodic retraining and remaining closely tied to the evolving distribution of OE retrievals.

Similar to Fig. 3, the ML predictions reproduce the spatial structure of total column CO with high fidelity, including regions of enhanced concentrations associated with wildfire activity over North America. Differences between ML predictions and colocated OE retrievals remain small and spatially unstructured. The median difference is < 0.13%, and the majority of samples (again, within the 5<sup>th</sup> and 95<sup>th</sup> percentiles) has ML CO concentrations within  $\pm 3.00\%$  of the OE results.

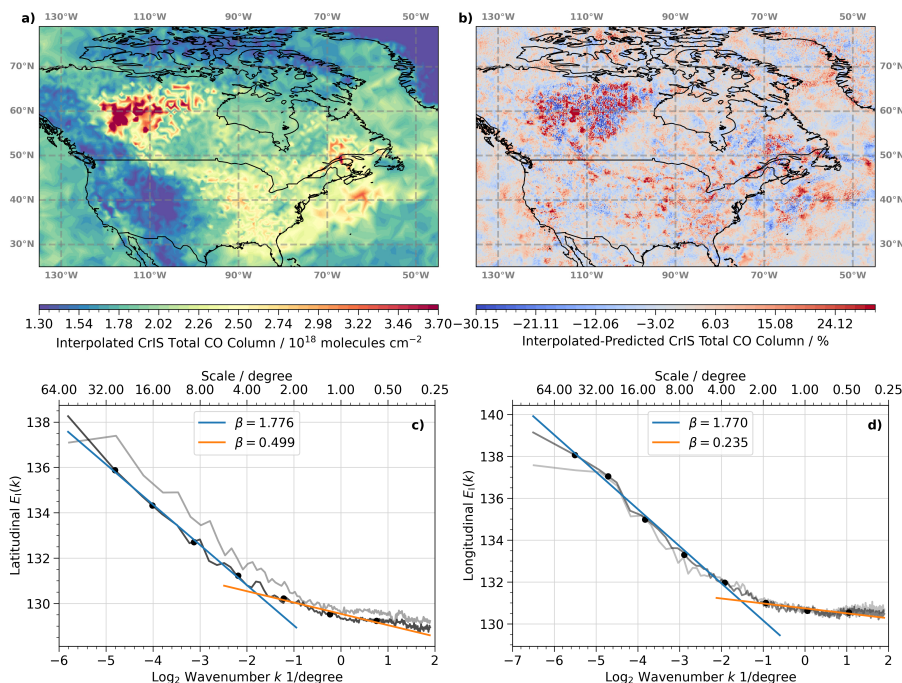
In addition to column concentrations, Fig. 4 demonstrates that the ML model accurately reproduces key retrieval diagnostics. The column AK exhibit a maximum at 383 hPa and is shown in panels b, d, and f, showing strong agreement in both spatial structure and magnitude. The median difference is < 0.67%, and the majority of ML predictions lie within  $\pm 8.50\%$  of the OE results. The global mean AK profiles (Fig. 4g) are nearly identical between OE and ML, with differences well within the natural variability of the OE retrieval. Median differences over all vertical levels are within 0.001 and 90% of predictions are within 0.028 of the true AK value at any level. This translates to 1.5% and 10%, respectively, at levels where the AK is noticeably different from zero, i.e., in the troposphere above the surface.

Likewise, differences in retrieval errors (Fig. 4h) and degrees of freedom (Fig. 4i) are centered near zero and exhibit narrow distributions, with full-width-at-half-maximum values of  $0.02 \cdot 10^{18}$  molec. cm<sup>-2</sup> and 0.032, respectively. This behavior indicates that the ML–predicted diagnostics are statistically consistent with those derived from OE, within the intrinsic variability of the retrieval.

Together, these results demonstrate that the ML framework not only reproduces the retrieved state, but also captures the associated sensitivity and uncertainty characteristics of the OE solution for unseen atmospheric conditions, while resolving



**Figure 4.** (a), (c), and (e) Similar to Fig. 3(a), (c), and (e), but showing global total column CO for 08 June 2025. (b), (d), and (f) Similar to (a), (c), and (e), but for the column averaging kernel (AK) at 383 hPa. (g) Global mean column AK profile as a function of pressure, showing OE results (blue), associated variability ( $\pm 1\sigma$ , shaded), and ML predictions (orange). (h) Probability density functions of the difference between ML and OE total column retrieval errors (blue) and tropospheric column retrieval errors (orange). (i) Probability density function of the difference between ML and OE degrees of freedom (DoF). This scene lies outside the training period.



**Figure 5.** (a) Interpolated TROPES total column CO on 10 June 2023. (b) Difference between interpolated and predicted CO. (c)–(d) Average power spectral density  $E_1(k)$  (black) as a function of wavenumber  $k$  for CO in latitudinal and longitudinal direction;  $E_1(k)$  for radiances at  $2183.125 \text{ cm}^{-1}$  are shown in gray. Blue and orange lines indicate linear fits through different regions of  $E_1(k)$ .

finer spatial structures beyond the native OE sampling. These characteristics suggest that the predicted diagnostics retain the key properties required for downstream applications such as data assimilation, although a full assessment within an assimilation framework is beyond the scope of this study.

## 235 4.2 The added value from CO at higher spatial resolution

A key question is whether the ML product captures physically meaningful CO variability below the nominal  $0.80^\circ$  TROPES retrieval resolution, or whether it primarily behaves as a spatial interpolation of the OE field. While interpolation can reconstruct smooth fields between observations, it cannot introduce new information at unresolved spatial scales.

To investigate this, we employ two complementary approaches: (i) comparing the ML-predicted CO fields with linearly interpolated TROPES CO retrievals, and (ii) analyzing the spatial power spectral densities  $E_1(k)$  to identify scale-dependent variability, particularly in the sub- $0.80^\circ$  domain. Together, these approaches allow us to assess whether interpolation is sufficient to capture the underlying structure, or whether additional variability persists at smaller scales that requires the higher-resolution information provided by the ML model.

Figure 5a shows the interpolated OE CO retrievals. This field appears significantly smoother than the ML predictions (Fig. 3c), especially for the region of enhanced CO over Western Canada and the Northeastern United States. The difference

between the interpolated and predicted CO concentrations is illustrated in Fig. 5b, where blue and red colors indicate underestimation and overestimation by the interpolation, respectively. Deviations are centered around 0.50% but can exceed  $\pm 30\%$ , especially in areas of enhanced CO. Maximum differences increase further, to  $\pm 48\%$  and  $\pm 58\%$ , when using nearest-neighbor or cubic spline interpolation, respectively.

250 While these differences highlight limitations of interpolation in regions of enhanced CO, a more general assessment of spatial variability across scales requires a spectral analysis. We therefore calculate power spectral densities  $E_1(k)$ , which describe how variance in a spatial signal is distributed across different wavenumbers ( $k$ ). Since this analysis requires data on a regular grid, the ML-predicted CO concentrations are first interpolated onto a grid with constant spacing. As in the previous section, we focus on the total column CO field over North America on 10 June 2023, gridded at a resolution of  $0.80^\circ/6 \approx 0.133^\circ$  in  
 255 both latitude and longitude. Because the CrIS L1B radiances and corresponding ML predictions are provided on a similar but irregular grid, nearest-neighbor interpolation is used to retain most of the native variability; for comparison, linear and cubic spline interpolations are also evaluated.

Many geophysical fields exhibit scale-invariant behavior over a large range of wavenumbers, with  $E_1(k)$  following a power law:

$$260 \quad E_1(k) \sim k^{-\beta}. \tag{1}$$

Sudden changes in the slope  $\beta$ , so-called scale breaks, indicate changes in the physical processes governing variability. Such breaks have been reported in cloud-reflected radiances (e.g., Davis et al., 1997), paleotemperature records (e.g., Nilsen et al., 2016), and climate variability (e.g., Franzke et al., 2020). We compute  $E_1(k)$  as the squared amplitude of the Fourier-transformed CO predictions in both latitudinal and longitudinal directions.

265 Figures 5c–d present  $E_1(k)$  averaged over all grid points in latitude and longitude, respectively. A scale break is observed at  $k \approx -1.70$ , corresponding to spatial scales of  $\approx 3.0\text{--}3.5^\circ$ , in both directions. Linear fits before and after the break, shown in blue and orange, were computed using the octave binning method reported in Davis et al. (1996), which mitigates noise and limits energy accumulation at small scales. The binned  $E_1(k)$  values are plotted as black dots. At the scale break, the slope in latitude (longitude) flattens from  $\beta \approx 1.77$  (1.77) to  $\beta \approx 0.49$  (0.23), indicating a relative enhancement of small-scale CO  
 270 variability. This transition is consistent with the de-correlation length scale of CO (not shown) and indicates a transition to enhanced small-scale variability. A detailed attribution of the underlying processes is beyond the scope of this study. Notably, no secondary break is observed at smaller scales, and no steeper decline in  $E_1(k)$  is found below the operational TROPES retrieval resolution, which would imply reduced variability and a smoother distribution. Instead, the CO field remains highly variable down to the Nyquist limit of  $2 \cdot 0.80^\circ/6 \approx 0.267^\circ$  (derived from the effective sampling resolution).

275 Importantly, the observed scale break is neither an artifact of the retrieval nor dependent on the interpolation scheme. Applying the same analysis to CO-sensitive radiances in the spectral microwindow used in the OE retrieval (gray line in Figs. 5c–d) yields similar results: comparable scale breaks at  $k \approx -2$  and consistent  $\beta$  values. Changing the interpolation scheme from nearest neighbor to linear or cubic spline has minimal effect on the location of the break, though  $\beta$  values flatten slightly,

reflecting increased variability across all scales. These minimal changes are not surprising, since the ML data are available at  
280 a very high spatial resolution (albeit on an irregular spatial grid).

For comparison, the power spectral density of linearly interpolated OE fields exhibits a similar large-scale slope ( $\beta \approx 1.70$ ) but a much steeper decay of variability toward smaller spatial scales ( $\beta \approx 3.28$ ), reflecting the smoothing inherent in interpolation. While the apparent scale break shifts slightly, its exact location is sensitive to the fitting procedure and binning choices and is therefore not interpreted further.

285 In summary, the results in this section demonstrate that significant variability in total CO concentrations exists at scales below  $\approx 4^\circ$ , and importantly, below the nominal  $0.80^\circ$  resolution of the TROPES retrievals. While interpolated OE fields exhibit a strong suppression of variability at these smaller scales, the ML product retains substantial structure down to the Nyquist limit. This indicates that the ML model captures physically meaningful sub-resolution variability that is not recovered by interpolation alone.

### 290 4.3 Computational costs

A key advantage of applying machine learning models in inference mode is their computational efficiency (Werner et al., 2023). As expected, the ML model is able to process a full day of CrIS radiance observations with remarkable speed. For 10 June 2023, the OE algorithm generated 44,192 CO column retrievals in  $\approx 160$  minutes. In contrast, the ML model predicted CO concentrations and associated diagnostics for 2,916,000 columns in just  $\approx 6$  minutes.

295 This performance difference is even more striking when considering the computational resources used. The OE algorithm was run on 60 compute nodes utilizing a total of 480 CPU cores, while the ML model required only a single compute node with 8 CPU cores. Additionally, the prediction success rate was higher: 98.4% for the ML model (based on a conservative outlier flag) compared to 90.69% for the OE retrievals.

The superior computational performance of the ML model ensures that every individual CrIS sample can be processed  
300 efficiently, enabling predictions for any species included in the TROPES retrieval framework. Moreover, this efficiency opens the door to near-real-time applications and provides a practical means to use the ML outputs to help constrain or enhance OE retrievals (see the discussion in Sect. 5).

Note, however, that the OE algorithm produces a vertical profile and an associated AK matrix whereas TROPES-HYREF currently only predicts the derived column.

## 305 5 Conclusions

This study presents a hybrid fusion of physics-based optimal estimation (OE) retrievals and machine learning (ML) to generate global, high-resolution carbon monoxide (CO) concentrations and associated diagnostics from CrIS observations. Our approach leverages the strengths of the TROPES OE retrievals, namely accuracy, physical consistency, and interpretability, while using an artificial neural network to overcome their main limitation: sparse spatial sampling due to high computational  
310 costs and strict quality filtering. This enables us to increase the fraction of processed CrIS observations from  $\approx 1\%$  to 100%.

The trained ML model within this Machine Learning–Optimal Estimation (TROPESS–HYREF) framework reproduces TROPESS CO column retrievals with high accuracy, achieving correlations exceeding 0.99 and low absolute biases  $< 0.1\%$  across both test and validation data sets. Importantly, TROPESS–HYREF predicts not only column concentrations, but also associated diagnostics, including column averaging kernels, degrees of freedom (DoF), and retrieval errors. The close agreement between predicted and OE-derived quantities demonstrates that the ML model successfully emulates both the retrieved state and its associated sensitivity and uncertainty characteristics. We emphasize that the ML component is designed to reproduce and extend the OE retrieval, rather than to surpass its physical accuracy, by providing full spatial coverage and enhanced resolution consistent with the underlying solution.

Using representative example scenes, we demonstrate that the TROPESS–HYREF predictions reproduce and extend fine-scale spatial structures consistent with the OE retrieval and outperform standard interpolation methods, particularly in regions with elevated CO due to wildfire emissions. A scale analysis reveals significant spatial variability in the CO fields below  $3.5^\circ$  and, more importantly, below the OE retrieval’s native  $0.80^\circ$  resolution, indicating that the ML predictions resolve meaningful sub-retrieval-scale features. Notably, variability persists down to the Nyquist sampling limit imposed by the CrIS observation footprint.

In terms of computational performance, TROPESS–HYREF processes a full day of CrIS observations more than 25 times faster than the OE algorithm, despite producing predictions for over 65 times more observations (i.e., 1,625 times faster). The high success rate of the ML inference ( $\approx 98.4\%$ ) compared to the OE retrieval ( $\approx 90.7\%$ ) further ensures consistent, global data availability.

By providing retrieval-like products at full coverage and enhanced resolution, this work bridges the gap between physically constrained atmospheric retrievals and scalable machine learning predictions. The ML outputs are suitable for downstream applications, including data assimilation, model validation, and trend analysis, while maintaining consistency with OE-derived physical information. They also offer the potential to inform and constrain future retrieval efforts, for example by serving as prior states or as an additional quality flag, where large discrepancies between OE and ML results could highlight potential issues with individual samples. In the current implementation, the model is trained on approximately two years of TROPESS OE retrievals; however, ongoing work focuses on incorporating regular retraining using newly available OE data. This ensures that TROPESS–HYREF continuously adapts to evolving atmospheric conditions while maintaining consistency with the OE retrieval, reinforcing its role as a complementary extension rather than a stand-alone predictive model.

The methodology developed for CO can be readily extended to other trace gases retrieved by the TROPESS MUSES OE algorithm, including ammonia ( $\text{NH}_3$ ), ozone ( $\text{O}_3$ ), and methane ( $\text{CH}_4$ ), as well as to other satellite instruments (TES, AIRS, OMI, and TROPOMI) and multi-instrument configurations (CrIS+TROPOMI or AIRS+OMI) (Fu et al., 2018; Malina et al., 2024). In parallel, ongoing work explores the use of ML predictions as a first guess in the OE retrieval algorithm, which may accelerate convergence and reduce computational costs. This hybrid framework therefore complements, rather than replaces, the OE retrieval, enabling scalable, physically consistent, high-resolution atmospheric composition products.

*Code and data availability.* CrIS L1B radiances and the TROPES CO product files can be downloaded from GES DISC. A Zenodo repository (Werner et al., 2025) contains the HYREF CO model and all necessary Python routines, as well as a Jupyter notebook with step-by-step instructions, so interested parties can produce their own CO predictions. This repository also includes Jupyter notebooks, Python routines, and ancillary data sets to reproduce each figure in the manuscript.

*Author contributions.* FW, KWB, SL, JLL, VHP, and JLM have shaped the concept of this study and refined the approach during extensive discussions. FW, SL, and JLM implemented the ML algorithm into the current TROPES algorithm pipeline. FW carried out the data analysis and prepared the figures for the manuscript. FW wrote the initial draft of the manuscript, which was subsequently refined by all authors.

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

*Acknowledgements.* Government sponsorship acknowledged. Work at the Jet Propulsion Laboratory, California Institute of Technology, was carried out under contract with the National Aeronautics and Space Administration (80NM0018D0004).

## References

- 355 Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, arXiv [preprint], arxiv:1603.04467v2, 2016.
- 360 Aumann, H., Chahine, M., Gautier, C., Goldberg, M., Kalnay, E., McMillin, L., Revercomb, H., Rosenkranz, P., Smith, W., Staelin, D., Strow, L., and Susskind, J.: AIRS/AMSU/HSB on the Aqua mission: design, science objectives, data products, and processing systems, *IEEE Trans. Geosci. Remote Sens.*, 41, 253–264, <https://doi.org/10.1109/TGRS.2002.808356>, 2003.
- Beer, R., Glavich, T. A., and Rider, D. M.: Tropospheric emission spectrometer for the Earth Observing System’s Aura satellite, *Appl. Opt.*, 40, 2356–2367, <https://doi.org/10.1364/AO.40.002356>, 2001.
- 365 Bowman, K. and Henze, D. K.: Attribution of direct ozone radiative forcing to spatially resolved emissions, *Geophysical Research Letters*, 39, <https://doi.org/https://doi.org/10.1029/2012GL053274>, 2012.
- Bowman, K., Rodgers, C., Kulawik, S., Worden, J., Sarkissian, E., Osterman, G., Steck, T., Lou, M., Eldering, A., Shephard, M., Worden, H., Lampel, M., Clough, S., Brown, P., Rinsland, C., Gunson, M., and Beer, R.: Tropospheric emission spectrometer: retrieval method and error analysis, *IEEE Transactions on Geoscience and Remote Sensing*, 44, 1297–1307, <https://doi.org/10.1109/TGRS.2006.871234>,  
370 2006.
- Bowman, K. W.: TROPESSE CrIS-JPSS1 L2 Carbon Monoxide for Forward Stream, Summary Product V1, <https://doi.org/10.5067/JL1HT3NGEAW3>, 2021.
- Buchholz, R. R., Hammerling, D., Worden, H. M., Deeter, M. N., Emmons, L. K., Edwards, D. P., and Monks, S. A.: Links Between Carbon Monoxide and Climate Indices for the Southern Hemisphere and Tropical Fire Regions, *Journal of Geophysical Research: Atmospheres*,  
375 123, 9786–9800, <https://doi.org/https://doi.org/10.1029/2018JD028438>, 2018.
- Buchholz, R. R., Worden, H. M., Park, M., Francis, G., Deeter, M. N., Edwards, D. P., Emmons, L. K., Gaubert, B., Gille, J., Martínez-Alonso, S., Tang, W., Kumar, R., Drummond, J. R., Clerbaux, C., George, M., Coheur, P.-F., Hurtmans, D., Bowman, K. W., Luo, M., Payne, V. H., Worden, J. R., Chin, M., Levy, R. C., Warner, J., Wei, Z., and Kulawik, S. S.: Air pollution trends measured from Terra: CO and AOD over industrial, fire-prone, and background regions, *Remote Sensing of Environment*, 256, 112 275,  
380 <https://doi.org/https://doi.org/10.1016/j.rse.2020.112275>, 2021.
- Byrne, B., Liu, J., Lee, M., Yin, Y., Bowman, K. W., Miyazaki, K., Norton, A. J., Joiner, J., Pollard, D. F., Griffith, D. W. T., Velazco, V. A., Deutscher, N. M., Jones, N. B., and Paton-Walsh, C.: The Carbon Cycle of Southeast Australia During 2019–2020: Drought, Fires, and Subsequent Recovery, *AGU Advances*, 2, e2021AV000 469, <https://doi.org/https://doi.org/10.1029/2021AV000469>, e2021AV000469  
2021AV000469, 2021.
- 385 Byrne, B., Liu, J., Bowman, K. W., Pascolini-Campbell, M., Chatterjee, A., Pandey, S., Miyazaki, K., van der Werf, G. R., Wunch, D., Wennberg, P. O., Roehl, C. M., and Sinha, S.: Carbon emissions from the 2023 Canadian wildfires, *Nature*, 633, 835–839, <https://doi.org/10.1038/s41586-024-07878-z>, 2024.
- Chollet, F. et al.: Keras, <https://keras.io>, 2015.

- Clerbaux, C., Hadji-Lazaro, J., Payan, S., Camy-Peyret, C., Wang, J., Edwards, D. P., and Luo, M.: Retrieval of CO from nadir remote-sensing measurements in the infrared by use of four different inversion algorithms, *Appl. Opt.*, 41, 7068–7078, <https://doi.org/10.1364/AO.41.007068>, 2002.
- Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Herbin, H., Hurtmans, D., Pommier, M., Razavi, A., Turquety, S., Wespes, C., and Coheur, P.-F.: Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder, *Atmospheric Chemistry and Physics*, 9, 6041–6054, <https://doi.org/10.5194/acp-9-6041-2009>, 2009.
- 395 Davis, A., Marshak, A., Wiscombe, W., and Cahalan, R.: Scale invariance of liquid water distributions in marine stratocumulus. Part I: Spectral properties and stationarity issues, *J. Atmos. Sci.*, 53, 1538–1558, [https://doi.org/10.1175/1520-0469\(1996\)053<1538:SIOLWD>2.0.CO;2](https://doi.org/10.1175/1520-0469(1996)053<1538:SIOLWD>2.0.CO;2), 1996.
- Davis, A., Marshak, A., Cahalan, R., and Wiscombe, W.: The Landsat scale break in stratocumulus as a three-dimensional radiative transfer effect: Implications for cloud remote sensing, *J. Atmos. Sci.*, 54, 241–260, 1997.
- 400 Drummond, J. R., Zou, J., Nichitiu, F., Kar, J., Deschambaut, R., and Hackett, J.: A review of 9-year performance and operation of the MOPITT instrument, *Advances in Space Research*, 45, 760–774, <https://doi.org/https://doi.org/10.1016/j.asr.2009.11.019>, 2010.
- Edwards, D. P., Emmons, L. K., Hauglustaine, D. A., Chu, D. A., Gille, J. C., Kaufman, Y. J., Pétron, G., Yurganov, L. N., Giglio, L., Deeter, M. N., Yudin, V., Ziskin, D. C., Warner, J., Lamarque, J.-F., Francis, G. L., Ho, S. P., Mao, D., Chen, J., Grechko, E. I., and Drummond, J. R.: Observations of carbon monoxide and aerosols from the Terra satellite: Northern Hemisphere variability, *Journal of Geophysical Research: Atmospheres*, 109, <https://doi.org/https://doi.org/10.1029/2004JD004727>, 2004.
- 405 Field, R. D., Luo, M., Kim, D., Del Genio, A. D., Voulgarakis, A., and Worden, J.: Sensitivity of simulated tropospheric CO to subgrid physics parameterization: A case study of Indonesian biomass burning emissions in 2006, *Journal of Geophysical Research: Atmospheres*, 120, 11,743–11,759, <https://doi.org/https://doi.org/10.1002/2015JD023402>, 2015.
- Field, R. D., Luo, M., Fromm, M., Voulgarakis, A., Mangeon, S., and Worden, J.: Simulating the Black Saturday 2009 smoke plume with an interactive composition-climate model: Sensitivity to emissions amount, timing, and injection height, *Journal of Geophysical Research: Atmospheres*, 121, 4296–4316, <https://doi.org/https://doi.org/10.1002/2015JD024343>, 2016.
- Franzke, C. L. E., Barbosa, S., Blender, R., Fredriksen, H.-B., Laepple, T., Lambert, F., Nilsen, T., Rypdal, K., Rypdal, M., Scotto, M. G., Vannitsem, S., Watkins, N. W., Yang, L., and Yuan, N.: The Structure of Climate Variability Across Scales, *Reviews of Geophysics*, 58, e2019RG000657, <https://doi.org/https://doi.org/10.1029/2019RG000657>, e2019RG000657 10.1029/2019RG000657, 2020.
- 415 Fu, D., Bowman, K. W., Worden, H. M., Natraj, V., Worden, J. R., Yu, S., Veeckind, P., Aben, I., Landgraf, J., Strow, L., and Han, Y.: High-resolution tropospheric carbon monoxide profiles retrieved from CrIS and TROPOMI, *Atmospheric Measurement Techniques*, 9, 2567–2579, <https://doi.org/10.5194/amt-9-2567-2016>, 2016.
- Fu, D., Kulawik, S. S., Miyazaki, K., Bowman, K. W., Worden, J. R., Eldering, A., Livesey, N. J., Teixeira, J., Irion, F. W., Herman, R. L., Osterman, G. B., Liu, X., Levelt, P. F., Thompson, A. M., and Luo, M.: Retrievals of tropospheric ozone profiles from the synergism of AIRS and OMI: methodology and validation, *Atmospheric Measurement Techniques*, 11, 5587–5605, <https://doi.org/10.5194/amt-11-5587-2018>, 2018.
- 420 Fu, D., Millet, D. B., Wells, K. C., Payne, V. H., Yu, S., Guenther, A., and Eldering, A.: Direct retrieval of isoprene from satellite-based infrared measurements, *Nature Communications*, 10, 3811, <https://doi.org/10.1038/s41467-019-11835-0>, 2019.
- Gaubert, B., Worden, H. M., Arellano, A. F. J., Emmons, L. K., Tilmes, S., Barré, J., Martinez Alonso, S., Vitt, F., Anderson, J. L., Alkemade, F., Houweling, S., and Edwards, D. P.: Chemical Feedback From Decreasing Carbon Monoxide Emissions, *Geophysical Research Letters*, 44, 9985–9995, <https://doi.org/https://doi.org/10.1002/2017GL074987>, 2017.
- 425

- Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning (Adaptive Computation and Machine Learning series), The MIT Press, Cambridge, MA, 2016.
- Grivas, G. and Chaloulakou, A.: Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece, *Atmospheric Environment*, 40, 1216–1229, <https://doi.org/10.1016/j.atmosenv.2005.10.036>, 2006.
- 430 Han, Y., Revercomb, H., Crompton, M., Gu, D., Johnson, D., Mooney, D., Scott, D., Strow, L., Bingham, G., Borg, L., Chen, Y., DeSlover, D., Esplin, M., Hagan, D., Jin, X., Knuteson, R., Motteler, H., Predina, J., Suwinski, L., Taylor, J., Tobin, D., Tremblay, D., Wang, C., Wang, L., Wang, L., and Zavyalov, V.: Suomi NPP CrIS measurements, sensor data record algorithm, calibration and validation activities, and record data quality, *Journal of Geophysical Research: Atmospheres*, 118, 12,734–12,748, <https://doi.org/https://doi.org/10.1002/2013JD020344>, 2013.
- 435 Holloway, T., Levy II, H., and Kasibhatla, P.: Global distribution of carbon monoxide, *Journal of Geophysical Research: Atmospheres*, 105, 12 123–12 147, <https://doi.org/https://doi.org/10.1029/1999JD901173>, 2000.
- IPCC: Climate Change 2023: Synthesis Report. A Report of the Intergovernmental Panel on Climate Change, IPCC, <https://doi.org/10.59327/IPCC/AR6-9789291691647>, 2023.
- 440 Jacob, D.: Introduction to Atmospheric Chemistry, vol. ISBN 9780691001852, Princeton University Press, 1999.
- Jain, P., Barber, Q. E., Taylor, S. W., Whitman, E., Castellanos Acuna, D., Boulanger, Y., Chavardès, R. D., Chen, J., Englefield, P., Flannigan, M., Girardin, M. P., Hanes, C. C., Little, J., Morrison, K., Skakun, R. S., Thompson, D. K., Wang, X., and Parisien, M.-A.: Drivers and Impacts of the Record-Breaking 2023 Wildfire Season in Canada, *Nature Communications*, 15, 6764, <https://doi.org/10.1038/s41467-024-51154-7>, 2024.
- 445 Jones, D. B. A., Bowman, K. W., Palmer, P. I., Worden, J. R., Jacob, D. J., Hoffman, R. N., Bey, I., and Yantosca, R. M.: Potential of observations from the Tropospheric Emission Spectrometer to constrain continental sources of carbon monoxide, *Journal of Geophysical Research: Atmospheres*, 108, <https://doi.org/https://doi.org/10.1029/2003JD003702>, 2003.
- Knobelspiesse, K., Tan, Q., Bruegge, C., Cairns, B., Chowdhary, J., van Diedenhoven, B., Diner, D., Ferrare, R., van Harten, G., Jovanovic, V., Ottaviani, M., Redemann, J., Seidel, F., and Sinclair, K.: Intercomparison of airborne multi-angle polarimeter observations from the Polarimeter Definition Experiment, *Appl. Opt.*, 58, 650–669, <https://doi.org/10.1364/AO.58.000650>, 2019.
- 450 Lelieveld, J., Gromov, S., Pozzer, A., and Taraborrelli, D.: Global tropospheric hydroxyl distribution, budget and reactivity, *Atmospheric Chemistry and Physics*, 16, 12 477–12 493, <https://doi.org/10.5194/acp-16-12477-2016>, 2016.
- Luo, M., Worden, H. M., Field, R. D., Tsigaridis, K., and Elsaesser, G. S.: TROPES-CrIS CO single-pixel vertical profiles: intercomparisons with MOPITT and model simulations for 2020 western US wildfires, *Atmospheric Measurement Techniques*, 17, 2611–2624, <https://doi.org/10.5194/amt-17-2611-2024>, 2024.
- 455 Malina, E., Bowman, K. W., Kantchev, V., Kuai, L., Kurosu, T. P., Miyazaki, K., Natraj, V., Osterman, G. B., Oyafuso, F., and Thill, M. D.: Joint spectral retrievals of ozone with Suomi NPP CrIS augmented by S5P/TROPOMI, *Atmospheric Measurement Techniques*, 17, 5341–5371, <https://doi.org/10.5194/amt-17-5341-2024>, 2024.
- Miller, D. J., Segal-Rozenhaimer, M., Knobelspiesse, K., Redemann, J., Cairns, B., Alexandrov, M., van Diedenhoven, B., and Wasilewski, A.: Low-level liquid cloud properties during ORACLES retrieved using airborne polarimetric measurements and a neural network algorithm, *Atmospheric Measurement Techniques*, 13, 3447–3470, <https://doi.org/10.5194/amt-13-3447-2020>, 2020.
- 460 Miyazaki, K., Eskes, H. J., and Sudo, K.: A tropospheric chemistry reanalysis for the years 2005–2012 based on an assimilation of OMI, MLS, TES, and MOPITT satellite data, *Atmospheric Chemistry and Physics*, 15, 8315–8348, <https://doi.org/10.5194/acp-15-8315-2015>, 2015.

- 465 Neyra-Nazarrett, O. A., Miyazaki, K., Bowman, K. W., and Saide, P. E.: An Assessment of TROPES CrIS and TROPOMI CO Retrievals and Their Synergies for the 2020 Western U.S. Wildfires, *Remote Sensing*, 17, <https://doi.org/10.3390/rs17111854>, 2025.
- Nilsen, T., Rypdal, K., and Fredriksen, H.-B.: Are there multiple scaling regimes in Holocene temperature records?, *Earth System Dynamics*, 7, 419–439, <https://doi.org/10.5194/esd-7-419-2016>, 2016.
- Noël, S., Reuter, M., Buchwitz, M., Borchardt, J., Hilker, M., Schneising, O., Bovensmann, H., Burrows, J. P., Di Noia, A., Parker, R. J.,  
470 Suto, H., Yoshida, Y., Buschmann, M., Deutscher, N. M., Feist, D. G., Griffith, D. W. T., Hase, F., Kivi, R., Liu, C., Morino, I., Notholt, J.,  
Oh, Y.-S., Ohyama, H., Petri, C., Pollard, D. F., Rettinger, M., Roehl, C., Rousogonous, C., Sha, M. K., Shiomi, K., Strong, K., Sussmann,  
R., Té, Y., Velazco, V. A., Vrekoussis, M., and Warneke, T.: Retrieval of greenhouse gases from GOSAT and GOSAT-2 using the FOCAL  
algorithm, *Atmospheric Measurement Techniques*, 15, 3401–3437, <https://doi.org/10.5194/amt-15-3401-2022>, 2022.
- Reed, R. and Marks, II, R. J.: *Neural Smoothing: Supervised Learning in Feedforward Artificial Neural Networks*, A Bradford Book, 1999.
- 475 Reichle Jr., H. G., Connors, V. S., Holland, J. A., Sherrill, R. T., Wallio, H. A., Casas, J. C., Condon, E. P., Gormsen, B. B., and Seiler, W.:  
The distribution of middle tropospheric carbon monoxide during early October 1984, *Journal of Geophysical Research: Atmospheres*, 95,  
9845–9856, <https://doi.org/https://doi.org/10.1029/JD095iD07p09845>, 1990.
- Rodgers, C.: *Inverse Methods for Atmospheric Sounding*, World Scientific Publishing Co., 2000.
- Saponaro, G., Kolmonen, P., Karhunen, J., Tamminen, J., and de Leeuw, G.: A neural network algorithm for cloud fraction estimation  
480 using NASA-Aura OMI VIS radiance measurements, *Atmospheric Measurement Techniques*, 6, 2301–2309, <https://doi.org/10.5194/amt-6-2301-2013>, 2013.
- Schultz, M. G., Akimoto, H., Bottenheim, J., Buchmann, B., Galbally, I. E., Gilge, S., Helmig, D., Koide, H., Lewis, A. C., Nov-  
elli, P. C., Plass-Dülmer, C., Ryerson, T. B., Steinbacher, M., Steinbrecher, R., Tarasova, O., Tørseth, K., Thouret, V., and Zell-  
weger, C.: The Global Atmosphere Watch reactive gases measurement network, *Elementa: Science of the Anthropocene*, 3, 000067,  
485 <https://doi.org/10.12952/journal.elementa.000067>, 2015.
- Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A., and Stadtler, S.: Can deep learning beat  
numerical weather prediction?, *Phil. Trans. R. Soc. A.*, 379, <https://doi.org/10.1098/rsta.2020.0097>, 2021.
- Strode, S. A., Worden, H. M., Damon, M., Douglass, A. R., Duncan, B. N., Emmons, L. K., Lamarque, J.-F., Manyin, M., Oman, L. D.,  
Rodriguez, J. M., Strahan, S. E., and Tilmes, S.: Interpreting space-based trends in carbon monoxide with multiple models, *Atmospheric*  
490 *Chemistry and Physics*, 16, 7285–7294, <https://doi.org/10.5194/acp-16-7285-2016>, 2016.
- Tyralis, H. and Papacharalampous, G.: A review of predictive uncertainty estimation with machine learning, *Artificial Intelligence Review*,  
57, 94, <https://doi.org/10.1007/s10462-023-10698-8>, 2024.
- UW-Madison Space Science and Engineering Center: Hank Revercomb; UMBC Atmospheric Spectroscopy Laboratory: Larrabee Strow:  
JPSS-1 CrIS Level 1B Full Spectral Resolution V2, <https://doi.org/10.5067/EETSCFBDBLX6>, 2018.
- 495 Veefkind, J., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H., de Haan, J., Kleipool, Q., van  
Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R.,  
Visser, H., and Levelt, P.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the at-  
mospheric composition for climate, air quality and ozone layer applications, *Remote Sensing of Environment*, 120, 70–83,  
<https://doi.org/https://doi.org/10.1016/j.rse.2011.09.027>, the Sentinel Missions - New Opportunities for Science, 2012.
- 500 von Clarmann, T. and Glatthor, N.: The application of mean averaging kernels to mean trace gas distributions, *Atmospheric Measurement*  
*Techniques*, 12, 5155–5160, <https://doi.org/10.5194/amt-12-5155-2019>, 2019.

- Werner, F., Schwartz, M. J., Livesey, N. J., Read, W. G., and Santee, M. L.: Extreme Outliers in Lower Stratospheric Water Vapor Over North America Observed by MLS: Relation to Overshooting Convection Diagnosed From Colocated Aqua-MODIS Data, *Geophysical Research Letters*, 47, e2020GL090131, <https://doi.org/10.1029/2020GL090131>, 2020.
- 505 Werner, F., Livesey, N. J., Schwartz, M. J., Read, W. G., Santee, M. L., and Wind, G.: Improved cloud detection for the Aura Microwave Limb Sounder (MLS): training an artificial neural network on colocated MLS and Aqua MODIS data, *Atmos. Meas. Tech.*, 14, 7749–7773, <https://doi.org/10.5194/amt-14-7749-2021>, 2021.
- Werner, F., Livesey, N. J., Millán, L. F., Read, W. G., Schwartz, M. J., Wagner, P. A., Daffer, W. H., Lambert, A., Tolstoff, S. N., and Santee, M. L.: Applying machine learning to improve the near-real-time products of the Aura Microwave Limb Sounder, *Atmospheric*  
510 *Measurement Techniques*, 16, 2733–2751, <https://doi.org/10.5194/amt-16-2733-2023>, 2023.
- Werner, F., Bowman, K. W., Lee, S., Laughner, J. L., Payne, V. H., and McDuffie, J. L.: Zenodo repository for A hybrid optimal estimation and machine learning approach to predict atmospheric composition, *Zenodo*, <https://doi.org/10.5281/zenodo.16968703>, 2025.
- Worden, H. M., Deeter, M. N., Frankenber, C., George, M., Nichitiu, F., Worden, J., Aben, I., Bowman, K. W., Clerbaux, C., Coheur, P. F., de Laat, A. T. J., Detweiler, R., Drummond, J. R., Edwards, D. P., Gille, J. C., Hurtmans, D., Luo, M., Martínez-Alonso, S., Massie,  
515 S., Pfister, G., and Warner, J. X.: Decadal record of satellite carbon monoxide observations, *Atmospheric Chemistry and Physics*, 13, 837–850, <https://doi.org/10.5194/acp-13-837-2013>, 2013.
- Worden, H. M., Francis, G. L., Kulawik, S. S., Bowman, K. W., Cady-Pereira, K., Fu, D., Hegarty, J. D., Kantchev, V., Luo, M., Payne, V. H., Worden, J. R., Commane, R., and McKain, K.: TROPES/CrIS carbon monoxide profile validation with NOAA GML and ATom in situ aircraft observations, *Atmospheric Measurement Techniques*, 15, 5383–5398, <https://doi.org/10.5194/amt-15-5383-2022>, 2022.
- 520 Zeng, Z.-C., Lee, L., Qi, C., Clarisse, L., and Van Damme, M.: Optimal estimation retrieval of tropospheric ammonia from the Geostationary Interferometric Infrared Sounder on board FengYun-4B, *Atmospheric Measurement Techniques*, 16, 3693–3713, <https://doi.org/10.5194/amt-16-3693-2023>, 2023.
- Zheng, B., Chevallier, F., Yin, Y., Ciais, P., Fortems-Cheiney, A., Deeter, M. N., Parker, R. J., Wang, Y., Worden, H. M., and Zhao, Y.: Global atmospheric carbon monoxide budget 2000–2017 inferred from multi-species atmospheric inversions, *Earth System Science Data*, 11,  
525 1411–1436, <https://doi.org/10.5194/essd-11-1411-2019>, 2019.