TCOM-CFC11 and TCOM-CFC12: A Gap-Free, Observationally Constrained Global Dataset of Stratospheric CFC-11 and CFC-12 Profiles (v2.0)

Sandip S. Dhomse[1,2] and Martyn P. Chipperfield[1,2]

[1]School of Earth and Environment, University of Leeds, Leeds, UK
[2]National Centre for Earth Observation, University of Leeds, Leeds, UK

Correspondence: Sandip Dhomse (s.s.dhomse@leeds.ac.uk)

Abstract. Understanding the long-term trends of ozone-depleting substances (ODSs), particularly CFC-11 ($CFCl_3$) and CFC-12 (CF_2Cl_2), is essential for evaluating the effectiveness of the Montreal Protocol. However, reliably estimating these trends is complicated by the inherent sparse spatial and temporal coverage of high-quality stratospheric observations, such as those from the Atmospheric Chemistry Experiment–Fourier Transform Spectrometer (ACE-FTS). To address this limitation, we have developed an innovative machine learning methodology to combine the strengths of sparse ACE-FTS observations with the continuous output of the TOMCAT global ~~Chemical Transport Model~~ chemical transport model (CTM).

We use XGBoost regression to constrain the TOMCAT tracers against co-located ACE-FTS measurements, thereby creating the TCOM (TOMCAT CTM and ~~occultation-measurement-based~~Occultation-Measurement-based) stratospheric profile datasets for CFC-11 and CFC-12. The resulting TCOM datasets described here (version 2.0) provide continuous, gap-free, global, daily vertical profiles from 2000 to 2024. A comprehensive evaluation confirms the method's effectiveness, showing the corrected TCOM data clustering significantly closer to the observations than the CTM and successfully ~~removing a systematic low bias~~ rectifies systematic biases, particularly the prominent high biases in the mid-to-high latitudes and low biases in the lower stratosphere present in TOMCAT-simulated CFC concentrations. Furthermore, interpretable machine learning analysis reveals that the XGBoost model primarily functions here as a "transport corrector", with dynamical features (like ~~Age-of-Air~~age-of-air, temperature, long-lived-tracers) being highly influential. This suggests that the dominant source of bias in the baseline TOMCAT simulation for the long-lived source gases considered here relates to its simulation of stratospheric circulation from meteorological reanalyses. These TCOM datasets are publicly available at https://doi.org/10.5281/zenodo.18145730 (Dhomse, 2026a) and https://doi.org/10.5281/zenodo.18147392 (Dhomse, 2026b), ~~and will provide~~ providing a valuable, observationally-constrained benchmark for refining chemical models and reducing uncertainties in ~~ODS trend analyses .~~ ODS trend analyses

1 Introduction

Stratospheric ozone depletion remains a critical environmental concern, as the ozone layer provides a vital shield against harmful ultraviolet (UV) radiation, protecting life on Earth. In response to the threat posed by anthropogenic gases, the 1987 Montreal Protocol on Substances that Deplete the Ozone Layer stands as a landmark international environmental agreement, widely recognised for its effectiveness in addressing this global challenge (WMO, 2011, 2015, 2018, 2022). Among the ODSs controlled by the Protocol, trichlorofluoromethane ($CFCl_3$ or CFC-11) and dichlorodifluoromethane (CF_2Cl_2 or CFC-12) are historically the major contributors to the stratospheric chlorine budget, collectively accounting for over 50% of ~~the stratospheric chlorine budget~~peak anthropogenic loading. Continuous, high-precision monitoring of the atmospheric concentrations and stratospheric distribution of these gases is essential to assess the Protocol's long-term effectiveness and track the ozone layer's projected recovery to pre-1980 levels (e.g. Eyring et al., 2007, 2010; Dhomse et al., 2018; Chipperfield and Bekki, 2024).

In addition to their role as ODSs, these halogenated substances are also extremely potent greenhouse gases with high Global Warming Potentials (GWPs) and have contributed significantly to positive radiative forcing of the climate system (Ramanathan et al., 1985). The successful phase-out of these compounds, mandated by the Montreal Protocol, has therefore provided a substantial co-benefit by avoiding further climate warming, underscoring the Protocol's powerful dual impact on ozone recovery and climate change mitigation (Velders et al., 2007). Furthermore, modelling studies such as Chipperfield et al. (2015) have quantified the benefits already achieved, demonstrating that without the Protocol, the Antarctic ozone hole would have been around 40% larger by 2013, and a deep Arctic ozone hole would have ~~occurred during the exceptionally cold winter of~~ already occurred during exceptionally cold winters like 2010/11.

Following the initial success of the Montreal Protocol, observations confirmed a decline in atmospheric ODS concentrations (e.g. Montzka et al., 1999). However, recent studies revealed that the atmospheric concentration of CFC-11 was not decreasing as rapidly as anticipated, suggesting potential emissions from renewed or unreported global ~~emissions~~ production (e.g., Montzka et al., 2018). Subsequent research linked these emissions primarily to undocumented production in Asia (e.g., Rigby et al., 2019). Modelling studies indicate that such unexpected emissions could delay the expected recovery of the stratospheric ozone layer (e.g. Dhomse et al., 2018) by approximately 5 to 6 years (Dhomse et al., 2019).

Additionally, there ~~are large~~ remain important uncertainties in the lifetimes of these CFC species, primarily caused by uncertainties in both emission estimates and loss rates. For example, using ACE-FTS measurements and assuming a CFC-11 lifetime of 45 years, Brown et al. (2013) estimated a CFC-12 lifetime of ∼113 years. Using various chemistry model simulations and steady-state conditions, these were revised to about ∼~~55~~ 56 and ∼~~94~~ 96 years, respectively, in Chipperfield et al. (2014). Recent modelling studies have revised these numbers even further (∼66 and ∼96 years in Lickley et al. (2021); ∼50 and ∼ 86 years in Bourguet et al. (2025)). Hence, accurate and continuous long-term data records of stratospheric CFC profiles are extremely important for refining ODS lifetime estimates, evaluating the Protocol's success, and identifying emerging threats.

For the past couple of decades, satellite instruments, such as the ACE-FTS, the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), and the High Resolution Dynamics Limb Sounder (HIRDLS) have provided valuable vertical profile measurements of CFC gases. However, their coverage is often limited in space and time due to observational constraints.

For example, the longest time series (~~March 2004 to present~~late February 2004 to present) is available from ACE-FTS (Bernath, 2002; Bernath et al., 2005). However, ACE-FTS uses the solar occultation technique; hence, it only provides ~ 30 profiles per day. In contrast, as a limb sounder, MIPAS provided about 1000 profiles per day (Fischer et al., 2008) but data are available only from March 2002 to April 2012. HIRDLS provided more than 5000 profiles per day (Gille et al., 2008; Khosravi et al., 2009), but data are available only ~~for~~ from February 2002 to March 2005. Such sparsity, along with inter-instrument biases and differences in measurement techniques complicate trend analysis and model evaluation (e.g. Tegtmeier et al., 2016; Millán et al., 2018; Hegglin et al., 2021). Looking toward future observational capabilities, the Advanced Limb Infrared Chemistry Experiment (ALICE) instrument on NASA's STRIVE (Stratosphere Troposphere Response using Infrared Vertically-resolved light Explorer) satellite mission is designed to provide critical continuity for these measurements. STRIVE will measure high-resolution vertical profiles of O_3 and other important trace gases, including CFC-11 and CFC-12 but will not be launched until the early 2030s. These observations will be vital for ongoing efforts to monitor and understand the recovery of the ozone layer and track the effectiveness of the Montreal Protocol.

Until now, the only readily available profile data for the evaluation of CFC chemistry in chemical models has been from the SPARC (Stratosphere–troposphere Processes and their Role in Climate) Data Initiative (e.g. Tegtmeier et al., 2013; Hegglin et al., 2021). These multi-instrument efforts were designed to establish a reference dataset for stratospheric composition. They provided the first comprehensive assessment and compilation of measurements from a suite of space-based limb sounders, to construct profile data from the upper troposphere to the lower mesosphere ($\sim$ 300–0.1 hPa). These compilations consolidate the original satellite data into standardised, vertically resolved, zonal monthly mean time series for various atmospheric constituents. Within the SPARC framework, CFC-11 and CFC-12 data files are created by compiling the output from multiple satellite missions, such as ACE-FTS (v3.6, Boone et al. (2013)), MIPAS (v20 and v22, Eckert et al. (2016)), and HIRDLS (v7, Khosravi et al. (2009)), onto a common latitude–pressure grid and a monthly time resolution. However, the SPARC initiative provided separate data files for each instrument covering different time periods, with some of these datasets having large gaps for certain latitude bins, making it difficult to use for evaluation of output from global chemistry models. Since then, no attempt has been made to harmonise these time series to construct long-term records. To address this gap, we have developed the TCOM data set (TOMCAT CTM and Occultation Measurement based), which provides vertical profiles without gaps. It offers daily, global, and observationally constrained CFC-11 (TCOM-CFC11) and CFC-12 (TCOM-CFC12) over a 25-year period (2000–2024).

This manuscript describes the construction of the TCOM-CFC11 and TCOM-CFC12 v2.0 datasets. We present the details of the input data, including the satellite measurements (ACE-FTS) and the model setup (TOMCAT CTM), in Section 2. The TCOM methodology, which follows the ~~version 1~~ v1.0 approach presented in Dhomse and Chipperfield (2023) but with several important updates, is detailed in Section 3. This is followed by a description of the data preprocessing steps in Section 4. Section 5 provides the evaluation of the newly constructed dataset, along with an analysis of the effect of various key input variables on the results. Finally, we provide ~~Summary and Conclusions~~ our summary and conclusions in Section 6.

2 Data

The generation of the TCOM dataset relies on the synergistic integration of two core, publicly available input sources: stratospheric profile observations from the ACE-FTS satellite instrument and global output from the TOMCAT ~~Chemical Transport Model~~ chemical transport model (CTM).

2.1 ACE-FTS Satellite Data

As noted ~~above~~in Section 1, the ACE-FTS instrument, aboard the SCISAT satellite, utilises the solar occultation technique to measure infrared solar absorption spectra. The instrument has been operational since ~~April~~ late February 2004 and provides high-quality vertical profiles globally (Bernath et al., 2005). However, due to its viewing geometry (limited to sunrise and sunset events), the measurements are spatially and temporally sparse, yielding approximately 30 profiles per day. The instrument operates with a high spectral resolution of $0.02\ \mathrm{cm}^{-1}$ across a broad spectral range spanning 750 to $4400\ \mathrm{cm}^{-1}$. Vertically, the profiles possess a high native resolution, with a sampling interval of approximately 2 to $6\ \mathrm{km}$ and a vertical field-of-view (FOV) of 3–$4\ \mathrm{km}$.

The ACE-FTS retrieval algorithm is described in detail by Boone et al. (2005, 2013, 2020, 2023). Briefly, the ~~ACE-FTS~~ retrieval uses a global-fit nonlinear least-squares approach, meaning it simultaneously fits calculated atmospheric spectra to the observed spectra across a range of altitudes and in selected narrow spectral regions (microwindows). This process involves an iterative procedure where a forward model calculates the expected atmospheric transmission spectra, and is adjusted until it optimally matches the actual measurements. A key characteristic of the ACE-FTS retrieval approach is that it does not employ optimal estimation or apply significant smoothing constraints~~;~~, hence, it does not rely heavily on a priori information. This methodology results in a high vertical resolution, and prior information has a negligible influence on the retrieved profiles except at the very top of the retrieval range where the atmosphere becomes optically thin.

For CFC-11, the retrieval uses 12 spectral windows starting from $829.03\ \mathrm{cm}^{-1}$ to $2979.50\ \mathrm{cm}^{-1}$ with the broadest microwindow centred around $845.50\ \mathrm{cm}^{-1}$ (with a $5.50\ \mathrm{cm}^{-1}$ width), covering the vertical range of 5–$28\ \mathrm{km}$. This retrieval requires careful correction for interfering gases, including HNO_3, H_2O, C_2H_6, and CO_2. For CFC-12, 15 spectral windows are used, and two broadest ones are centred at $921.50\ \mathrm{cm}^{-1}$ (covering 5–$28\ \mathrm{km}$) and ~~a second~~ at $1161.07\ \mathrm{cm}^{-1}$ (covering 15–$35\ \mathrm{km}$). The primary interfering species for CFC-12 retrievals include H_2O, CO_2, CH_4, N_2O, and O_3. The first version of the TCOM dataset (v1.0) utilised ACE-FTS v5.2 data. Here we detail the methodology of the updated product, TCOM v2.0, which incorporates the latest ACE-FTS v5.3 data product, which was released in February 2025. All ACE-FTS data are accessed via the ACE-FTS Data archive: https://databace.scisat.ca.

2.2 TOMCAT CTM

The TOMCAT model output serves as the background field, providing the spatial and temporal continuity required for the data gap-filling process. TOMCAT is an offline, global 3D CTM that incorporates a detailed stratospheric chemistry scheme (Chipperfield, 1999, 2006). The model's dynamics are driven here by the ERA5 reanalysis data (Hersbach et al., 2020), ensuring that the model includes our most updated knowledge about the dynamical state of the past atmosphere. TOMCAT has the capability to have variable vertical and horizontal resolution. Simulations used here are performed at horizontal resolution (spatial grid) of approximately $2.8° \times 2.8°$, corresponding to 64 latitude and 128 longitude gridpoints, with 32 vertical sigma-pressure levels, extending from the surface up to approximately 60 km. The model simulations are performed for a 25-year time period (following earlier spin-up), from 2000 to 2024, with daily global output fields saved at 1:30 PM local time.

Surface boundary conditions used in TOMCAT are based on the WMO (2022) scenario A1 for long-lived source gases, including CFCs, HFCs, CH_4, and N_2O (WMO, 2022). Other critical time-varying inputs include solar spectral irradiances (SSI) and Stratospheric Aerosol ~~Surface Area Density~~ surface area density (SAD) until December 2024 as described in ~~(Dhomse et al., 2015, 2016, 2022)~~Dhomse et al. (2015, 2016, 2022). Details of the construction of SSI (NRL ~~V2~~22) data are described in Coddington et al. (2016). After January 2018, the model uses SAD from Knepp et al. (2024).

3 Methodology

Following the approach outlined in Dhomse et al. (2021) (to construct a monthly mean zonal mean ozone profile dataset using Random Forest) and Dhomse and Chipperfield (2023) (which extended the methodology to daily datasets for CH_4 and N_2O), we employ supervised machine learning (ML) techniques to generate continuous, high-resolution vertical profiles of CFC-11 and CFC-12 volume mixing ratios (VMRs). The novel data–model methodology is specifically tailored for the generation of a long-term gap-free profile dataset for various stratospheric species. The core principle is to systematically constrain the simulated output of the CTM with co-located satellite observations. This integration is achieved through XGBoost regression (Chen and Guestrin, 2016), a powerful machine learning technique trained to correct the inherent biases in the CTM simulations, which are assumed to be a consequence of the parameters used in the chemical scheme (e.g., reaction rates, photolysis rates) as well as the model setup (e.g., horizontal/vertical resolution, chemical/dynamical time steps, forcing meteorology, as well as parameterisations used for various computationally expensive processes).

The construction of the TCOM data required a regression model that offered the optimal balance of predictive accuracy and robustness at different altitudes and latitude bands. A rigorous comparative analysis was performed across seven regression models, including traditional, regularised, and ensemble techniques. This systematic approach ensures the final selection is the optimal algorithm for the task (see Supplementary Figures S1 and S2).

Initially, we tested a simple linear regression (OLS), which served as a baseline. Following this, three regularised regression models were tested: Lasso (ℓ_1 regularisation, (Tibshirani, 1996)), Ridge (ℓ_2 regularisation, (Hoerl and Kennard, 1970)), and ElasticNet (Zou and Hastie, 2005). These models are particularly valuable in attribution-related studies (e.g., Li et al., 2022, 2023) because their penalty terms help to mitigate issues like multicollinearity and perform implicit feature selection~~, respectively~~. For instance, Lasso's strength lies in forcing some coefficients to exactly zero, simplifying the model, while Ridge's strength is stabilising estimates by shrinking all coefficients toward zero, which is effective when predictors are highly correlated. ElasticNet combines the strengths of both.

Finally, the performance of three ensemble models based on decision trees were analysed: Random Forest (RF) (Breiman, 2001), AdaBoost (Freund and Schapire, 1997), and XGBoost (eXtreme Gradient Boosting, (Chen and Guestrin, 2016)). These models, used in our previous studies (e.g. Dhomse et al., 2021; Dhomse and Chipperfield, 2023), are known for their ability to capture complex, non-linear relationships. While Random Forest reduces variance via averaging, AdaBoost sequentially corrects errors. XGBoost, however, provides an optimised and highly scalable gradient boosting framework with built-in regularisation, often yielding superior predictive accuracy.

Model performance was assessed using two complementary metrics: the RMSE and the coefficient of determination (R^2). The combined use of these metrics is ~~crucial~~ important for a comprehensive evaluation. The RMSE measures the average magnitude of the error in the model's predictions, expressed in the same units as the target variable:

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

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

Here, y_i represents the individual co-located ACE-FTS observed VMR values, which serve as the target variable for the regression, and $\hat{y}_i$ denotes the corresponding VMR values predicted by the XGBoost. N is the total number of observations in the test data set (30% of available data points; 70% are used for training).

A lower RMSE signifies higher accuracy. The metric is particularly important because, due to the squaring of residuals, it ~~heavily~~ penalises large errors or outliers, which are often significant in environmental data such as trace gas concentrations.

The R^2 quantifies the proportion of the variance in the dependent variable that is predictable from the independent variables:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{2}$$

where $\bar{y}$ is the arithmetic mean of the ACE-FTS observations.

A significant update in the v2.0 framework is that y now represents absolute mixing ratios rather than the model–observation differences used in version 1.0; this transition has been shown to improve regression performance and allows for more straightforward uncertainty quantification. The R^2 provides a standardised measure of the model's coefficient of determination, with values closer to ~~1~~ 1 indicating that the model explains most of the data's variability. While a high R^2 confirms a good overall fit, a low RMSE confirms that the actual prediction errors ~~that~~ are small in magnitude, ensuring both explanatory power and practical prediction utility.

After testing these parameters, XGBoost regression was ultimately selected for the TCOM data construction. ~~Supplementary figures (S1 to S2~~TCOM v1.0 (Dhomse and Chipperfield, 2023) primarily focussed on CH_4 and N_2O that was later extended for more species like HCl, HF, O_3, H_2O. Here we present results from updated methodology (e.g. improved pre-processing, larger feature matrix) and supplementary figures (S1 and S2) are updated versions of the analysis used for the previous TCOM ~~version 1.~~ v1.0. These figures clearly ~~demonstrates~~ demonstrate that, in terms of both performance metrics (R^2 and RMSE), the XGBoost model consistently shows the most effective performance for both CFCs. Similar superior performance patterns are observed for all the remaining latitude bands (not shown). This superior performance can be attributed to its advanced sequential boosting framework and effective built-in ℓ_1 and ℓ_2 regularisation, which together minimise the bias–variance trade-off. This selection ensures the use of the optimum algorithm for the TCOM data construction, ~~guaranteeing~~ helping to ensure maximum accuracy and robustness.

4 Data Preprocessing

Data preprocessing is a crucial and often time-consuming step in the machine learning workflow, as the quality of the input data directly dictates the performance and reliability of the final model output. ~~It~~ This preprocessing involves a series of transformations aimed at cleaning the raw data, handling missing values, managing outliers, and normalising or standardising features. By aligning, filtering, and structuring data into a suitable format, preprocessing helps to mitigate issues like noise, bias ~~,~~ and inconsistency. Furthermore, techniques such as feature engineering ~~allows~~ allow the model to better capture the underlying complexity and non-linear relationships within the dataset.

4.1 Data Filtering and Vertical Alignment

The initial preprocessing step involves ACE-FTS data filtering, specifically removing the measurements where the reported retrieval error exceeded a threshold of 200% including observations with negative mixing ratios. For optimised statistical analysis, the co-located TOMCAT profiles were interpolated to a uniform $1\ \mathrm{km}$ vertical resolution to align with the ACE-FTS retrieval grid.

To account for the distinct chemical and dynamical regimes across the globe, the training data were grouped into five distinct latitude bins:

1. Southern Hemisphere Polar (SHpol, $90°$S to $50°$S)
2. Southern Hemisphere Mid-latitudes (SHmid, $70°$S to $20°$S)
3. Tropics (Trop, $40°$S to $40°$N)
4. Northern Hemisphere Mid-latitudes (NHmid, $20°$N to $70°$N)
5. Northern Hemisphere Polar (NHpol, $50°$N to $90°$N)

This binning strategy ensures that the machine learning model is trained on local atmospheric characteristics. To maintain spatial continuity and to eliminate inhomogeneities, the corrected fields from overlapping latitude regions were averaged. Note that regression analysis was only performed when more than 2000 valid observations (with retrieval errors less than 200%) were available for a given altitude level. Generally, for tropospheric altitudes (below about 10 km at the poles and about 18 km in the tropics), the number of filtered data points available for XGBoost training is often limited (nearly 50% ~~less~~ fewer than stratospheric levels). ~~The somewhat relaxed filtering criteria ensures~~ Expanding error filters (200% vs 100% used in (Dhomse and Chipperfield, 2023)) ensures that data construction includes enough data points in the radiatively sensitive upper tropospheric levels.

4.2 Feature Engineering

The machine learning model's objective is to predict the observed ACE-FTS VMR (the target variable) by utilising information contained within the co-located TOMCAT output (the feature matrix). A key difference relative to TCOM ~~version 1~~ v1.0 (and methodology used in Dhomse and Chipperfield (2023)) is that here we use the absolute VMR values from ACE-FTS measurements as a target variable. Previously ~~version 1~~ v1.0 used the differences between the ACE-FTS satellite measurement and the related TOMCAT output variable. ~~These updates have improved the regression model performance and the calculation of uncertainty estimates now becomes straightforward.~~

In the previous version of this dataset, the feature matrix consisted of 32 variables, all of which were used during training. This included 12 terms to represent the seasonal cycle (monthly constants) and 20 tracers from the TOMCAT CTM, which encompassed key chemical species and dynamical variables: O_3, CH_4, N_2O, NO_2, HNO_3, HCl, HF, H_2O, the CFCs themselves (CFC-11 and CFC-12), related species (CFC-113, HCFC-141b, HCFC-142b, HFC-134a, COFCl, COF_2, CO), temperature (temp), age-of-air (AoA), and potential vorticity (PV). Three measurement-specific variables were also included: the measurement date (mea_date), latitude (lat), and the ACE-FTS retrieval error (err).

For this v2.0 dataset, a systematic feature reduction process was applied. An in-depth analysis indicated that the 12 seasonal cycle terms and the NO_2 term had a negligible impact on model performance, leading to their removal. Consequently, the initial feature matrix was reduced to 22 variables. To further mitigate multicollinearity and improve model efficiency for individual altitude levels, a refinement process was conducted. Highly correlated variables in the feature matrix (Pearson correlation coefficient $r > 0.9$) were analysed, and the variable exhibiting a lower correlation with the target VMR was excluded. Furthermore, features showing negligible correlation ($r < 0.05$) to the target were also removed, while ensuring a minimum of 10 features remained in the final matrix to maintain model robustness.

4.3 Dataset Construction and Uncertainty Quantification

The XGBoost regression model is trained independently for each 1 km altitude level and within each of the five latitude bins. The model's primary task is to identify the optimal combination of features that can most accurately predict the ACE-FTS observations. At each level and each latitude bin, 70% of the data points are used for the training and 30% are used for the testing. Once trained for a particular level and zonal bin, the resulting TCOM CFC VMR (D_{TCOM}) is constructed by predicting CFC ~~concentrations~~ VMRs on TOMCAT horizontal grid points (64 latitudes $\times$ 128 longitudes).

A critical component of this data product is the robust uncertainty estimate (σ_{TCOM}), calculated using an ensemble approach. The $\mathrm{XGBoost}$ model is retrained two additional times, using only the ACE-FTS observations corresponding to values below the $25\mathrm{th}$ or above the $75\mathrm{th}$ percentiles of the measured VMRs:

1. Upper Bound: Using only ACE-FTS observations larger than the $75\mathrm{th}$ percentile of the observations, creating $D_{\mathrm{TCOM},75}$.
2. Lower Bound: Using only ACE-FTS observations smaller than the $25\mathrm{th}$ percentile of the observations, creating $D_{\mathrm{TCOM},25}$.

The ~~final uncertainty is then defined~~ uncertainty estimate, σ_{TCOM}, is calculated as half the absolute difference between ~~these two resulting datasets:~~

$$\sigma_{\mathrm{TCOM}}(t,x,y,z) = \frac{1}{2} \times |D_{\mathrm{TCOM},75}(t,x,y,z) - D_{\mathrm{TCOM},25}(t,x,y,z)|$$

the upper and lower quantile predictions as:

$$\sigma_{\mathrm{TCOM}}(t,\lambda,\phi,z) = \frac{1}{2} |D_{\mathrm{TCOM},75}(t,\lambda,\phi,z) - D_{\mathrm{TCOM},25}(t,\lambda,\phi,z)| \tag{3}$$

where t is time, λ is longitude, ϕ is latitude, and z represents the vertical altitude level. The terms $D_{\mathrm{TCOM},75}$ and $D_{\mathrm{TCOM},25}$ denote the XGBoost model outputs retrained using only the ACE-FTS observations above the 75th and below the 25th percentiles, respectively.

Finally, the XGBoost model is trained independently for each 1 km altitude and latitude bin and is then used to construct data for all 128 longitudes for the given altitude and all the latitudes in a latitude bin for each day. The resulting 3D gap-free fields are generated by merging these reconstructed data to the global TOMCAT grid (64 latitudes × 128 longitudes). For overlapping latitudes, simple averaging is used. These updates have improved the regression model performance, and the calculation of uncertainty estimates now becomes straightforward.

5 Results and Discussion

Here we present analysis of two performance metrics, ~~in an attempt to~~ to demonstrate how and to explain why the model's predictive skill varies across different geographical and altitude regimes, with specific focus on the ~~unique~~ specific challenges presented by the tropical (Trop) latitude band.

5.1 Analysis of Performance Metrics

Figures 1 and 2 present the R^2 and RMSE values for all five latitude bands for CFC-11 and CFC-12, respectively, using the $\mathrm{XGBoost}$ regression model. For most latitude bands, the lowest possible altitude is about $5\ \mathrm{km}$, and the top altitude ranges from 25 km (for CFC-11) to about 30 km (for CFC-12). Overall, the R^2 values are generally better for the polar and mid-latitude bins (SHpol, SHmid, NHmid, NHpol) compared to the tropical latitude bin (Trop). The Trop bin exhibits the lowest R^2 values~~. However,~~ , however, the tropical band also shows the smallest RMSE values for tropospheric levels (below 18 km). This seemingly contradictory result can be attributed to three main factors affecting the prediction model's performance in the tropics:

1. Limited data availability: The SCISAT-1 satellite, which carries the ACE-FTS instrument, is in a high-inclination (74°) orbit. This orbital path naturally concentrates measurements at ~~higher~~ high and middle latitudes. Consequently, the tropical stratosphere has significantly fewer measurements ($\sim$21,000 measurements) compared to the polar/mid-latitude regions ($\sim$45,000). For tropospheric levels, the number of valid measurements for the tropical bands is even lower (~~$\sim$ 4,000~~$\sim$4,000), most likely due to the presence of thin layers of cirrus clouds near the tropical upper troposphere - lower stratosphere (UTLS) region, which impacts data retrieval quality.

2. High dynamical variability: The large dynamical variability associated with the younger AoA in the tropical stratosphere increases prediction complexity. This challenge is particularly pronounced in the troposphere and, to some extent, in the lower and middle ~~stratospheres~~stratosphere, making the relationship between the input features and the target variable less predictable.

3. Influence of chemical loss: The mid-to-high latitude distributions are controlled by a strong seasonal cycle in both horizontal transport from the tropics to high latitudes (e.g. Holton et al., 1995; Butchart and Scaife, 2001) and the downward transport of air that is CFC-poor (i.e., air that has ~~undergone significant photolysis loss~~experienced significant photolysis exposure) to lower levels. This dominant, annual transport mechanism in the mid-to-high latitudes results in a more robust fit for the XGBoost model, leading to higher R^2 values compared to the more complex tropical region.

Overall, the small RMSE values demonstrate that the absolute prediction errors remain low because the fundamental dynamical processes driving CFC VMRs are reasonably well-represented in TOMCAT. Conversely, the lower R^2 values are a result of the high dynamical variability and limited data availability in the tropics. In particular, at lower altitudes (below 17 km) ACE measures upper tropospheric VMRs, which makes it statistically more difficult for the model to explain a high proportion of the observed variance even when the absolute predictions are accurate.

5.1.1 CFC-11 Performance

For CFC-11, the performance of XGBoost in SHpol shows R^2 values exceeding 0.8 from approximately 12 km to ~~22 km~~21 km, with the best performance observed near 17 km (Figure 1). Above 21 km, R^2 values decrease rapidly. This is likely because CFC-11 ~~concentrations~~ VMRs are much smaller at these higher altitudes, leading to fewer observations and potentially larger retrieval errors. In the tropical latitude bin, the best performance ($R^2 \sim 0.9$) is observed near ~~21 km~~22 km, and R^2 values remain greater than 0.7 for nearly all altitudes above 10 km. As mentioned in Section 5.1, despite the lower R^2 values in the tropics, the RMSE errors remain below 15 pptv, suggesting that for this latitude band, the dynamical processes driving CFC-11

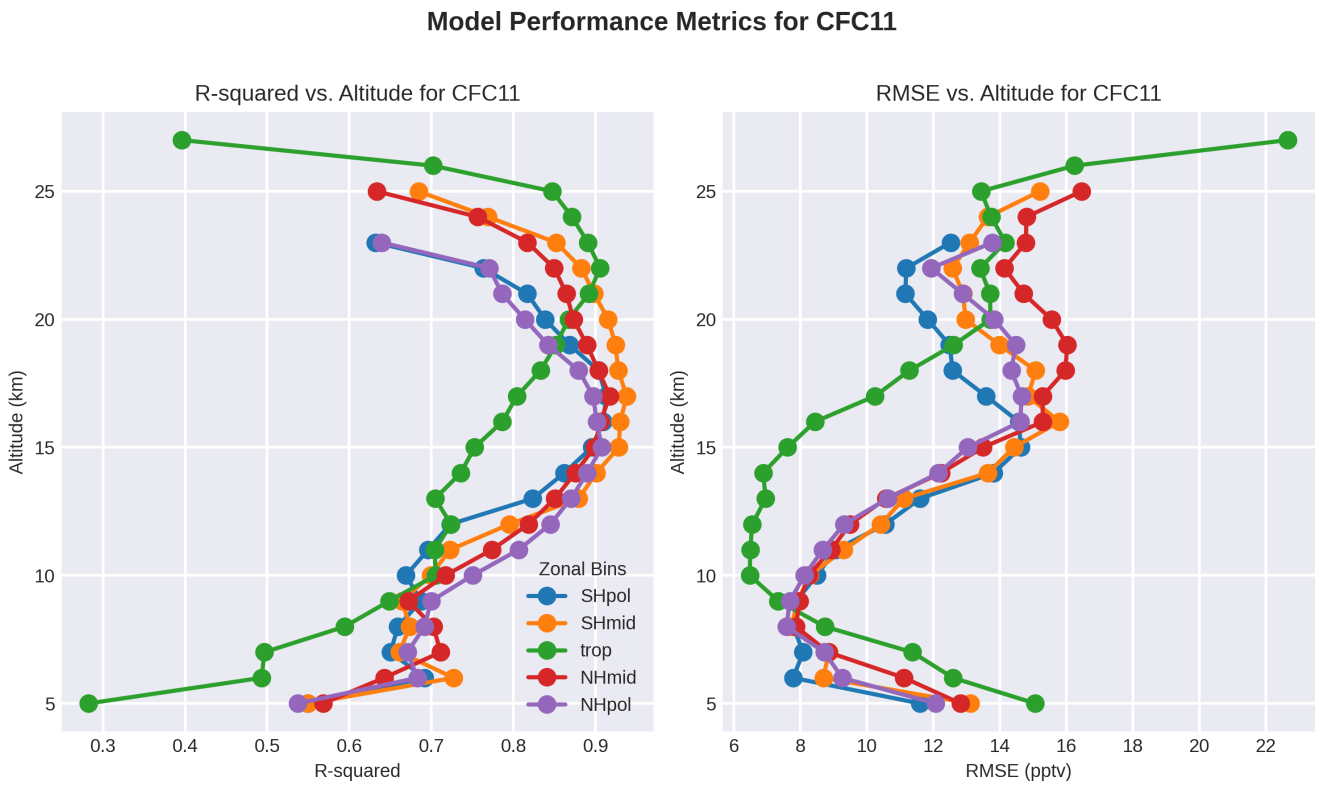

Figure 1. Coefficient of determination (R^2, left panel) and root mean squared error values (RMSE, right panel~~) values~~, in parts per trillion by volume (pptv)) for all 5-latitude bands from the XGBoost regression model.

~~concentrations~~ VMRs are reasonably well-represented in the TOMCAT model, which provides key features to the regression. The lowest RMSE values for all five latitude bands ~~is observed near~~ are observed near the UTLS region, indicating the ability of XGBoost to partially constrain dynamical variability in this ~~radiatively sensitive region .~~

region with large spatial and temporal variability that is caused by numerous competing transport, chemical, and mixing processes around the tropopause (e.g. Millán et al., 2023).

5.1.2 CFC-12 Performance

The performance metrics for CFC-12 (Figure 2) are generally better than those for CFC-11, with the tropical band being the exception. R^2 values for CFC-12 are close to 0.9 for altitudes ranging from 15 km to 25 km and remain above 0.8 for most stratospheric levels. The differences in R^2 values between CFC-11 and CFC-12 (particularly in the tropical band) are noteworthy. Since both species have long atmospheric lifetimes ($\sim$ ~~50 year~~ 56 years for CFC-11 and $\sim$ ~~90 year~~ 96 years for CFC-12, (Chipperfield et al., 2014)), they are co-emitted, and utilize nearly identical feature matrices in XGBoost~~indicating that~~ , so the discrepancies may be linked to differences in their loss processes. CFC-12 has a longer atmospheric lifetime and nearly double

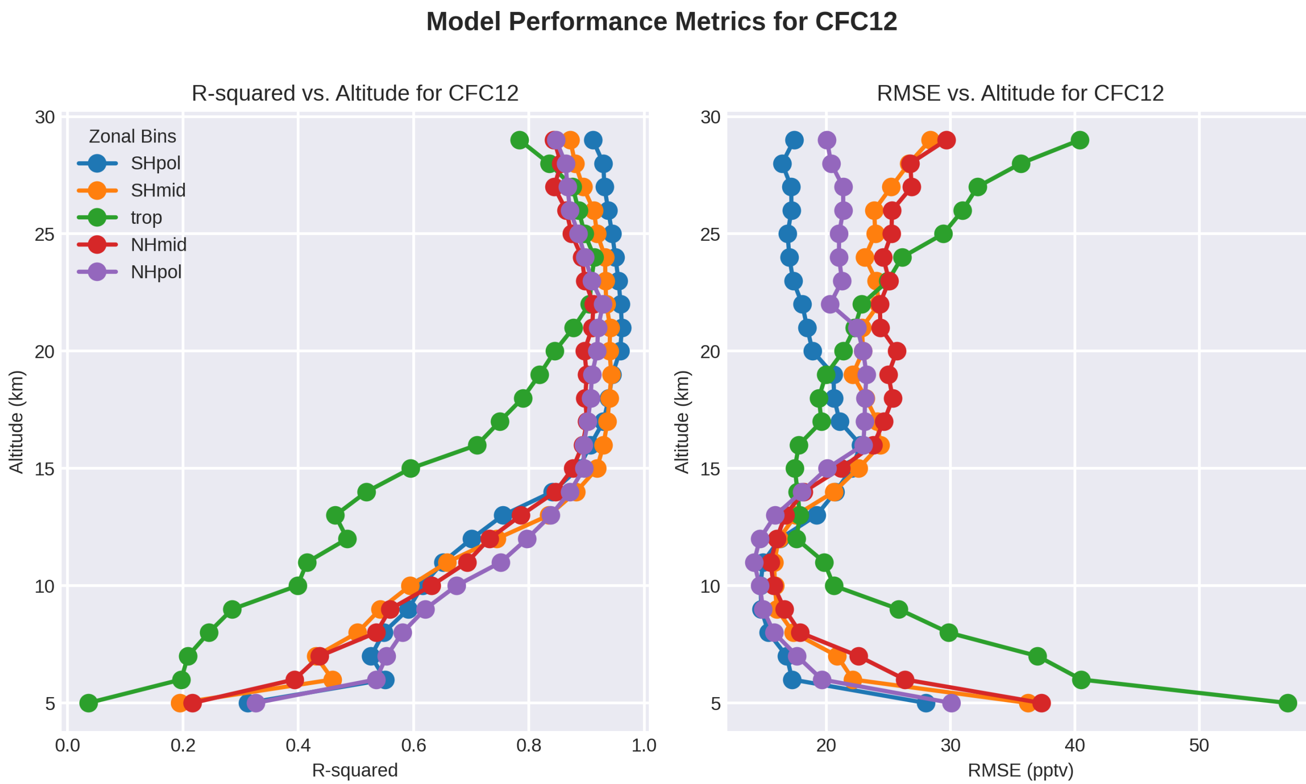

Figure 2. Same as Figure 1, but for CFC-12.

the concentration of CFC-11 ~~(Chipperfield et al., 2014; Hoffmann et al., 2014)~~(Hoffmann et al., 2014). Consequently, CFC-12 ~~concentrations~~ VMRs remain stable throughout most of the lower stratospheric levels. Furthermore, the CFC-12 model maintains high R^2 values over a more extended altitude range (up to 24 km) compared to CFC-11. This is most likely attributable to its ~~photolysis~~ photolytic loss occurring at higher altitudes, resulting in a more substantial and robust data set for training the regression model at mid-stratospheric levels (see Figure 2).

Additionally, vertical profiles of the retrieved CFC-11/CFC-12 mixing ratios (from ACE-FTS and the XGBoost model or TCOM) are compared with the original TOMCAT profiles in Supplementary Figures S3 and S4. The comparison, restricted to the 30% test data points, clearly shows that the TCOM profiles show very good agreement with the ACE-FTS observations across all zonal latitude bins and valid vertical levels. However, at mid-stratospheric levels (the top levels where ACE-FTS retrieval of individual species is possible), the XGBoost-derived profiles appear to shift towards the original TOMCAT profiles. This behaviour suggests a limitation in the XGBoost model's predictive ability at these altitudes, which is likely driven by the rapid decrease in CFC ~~concentrations~~VMRs . This decrease leads to very noisy CFC measurements (due to a poor signal-to-noise ratio in the satellite observations), resulting in a weakly trained and less constrained XGBoost model in the mid-stratosphere.

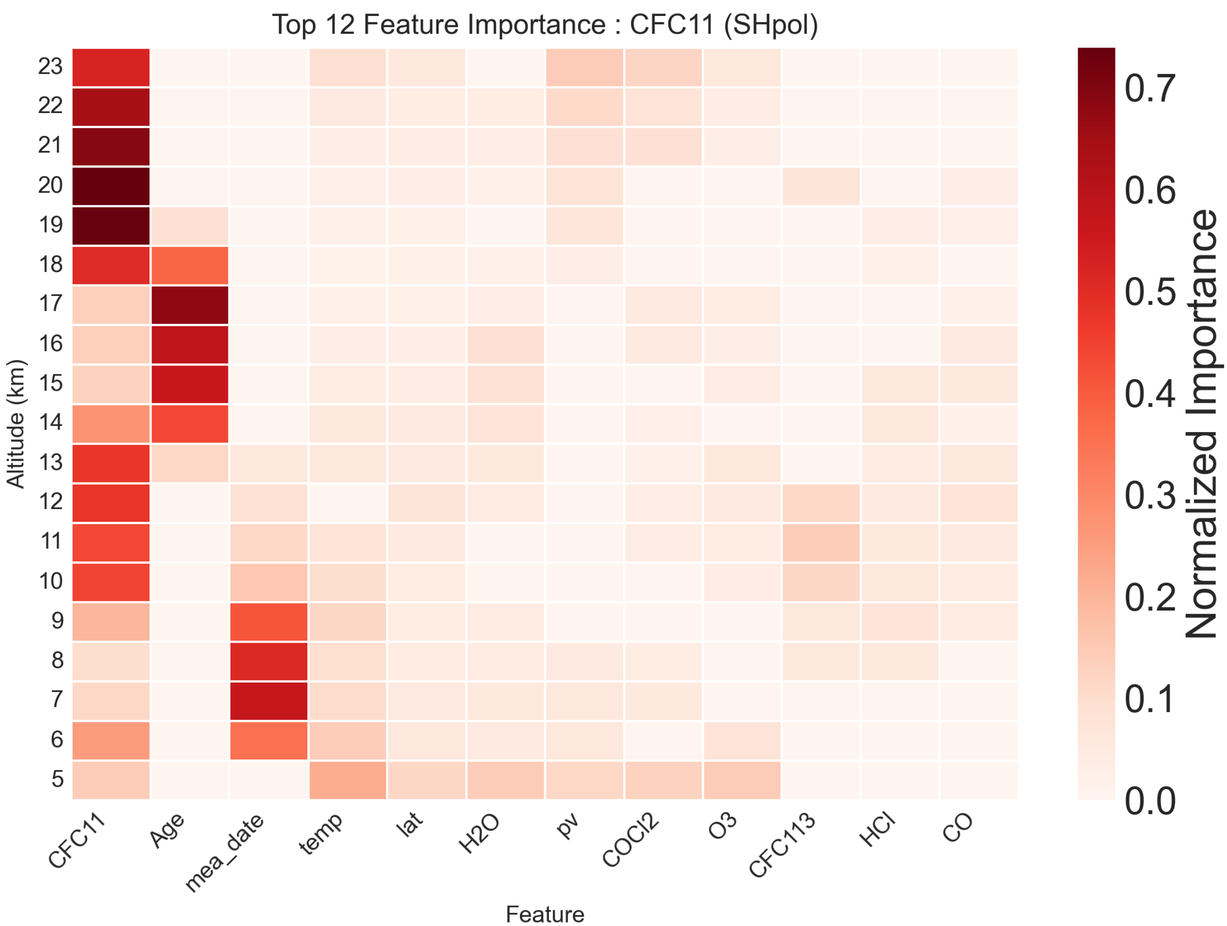

Figure 3. Normalised feature importance heatmap for CFC-11 in the Southern Hemisphere polar (SHpol) latitude bin, derived from the XGBoost regression model. The y-axis represents altitude (in km), and the x-axis lists the 12 most important input features. The colour intensity (red scale) indicates the normalised relative importance of each feature at a specific altitude level for predicting the ACE-FTS CFC-11 VMR.

5.1.3 Feature Importance Across Altitudes

The feature importance analysis, shown as heatmaps for the SHpol band (Figures 3 and 4), quantifies the overall influence of various variables or features on the XGBoost~~predicted~~ predicted CFC-11 and CFC-12 VMRs. Feature importances for other latitude bands are shown in Supplementary Figures S5 to S12. Overall, for both CFC-11 and CFC-12 XGBoost selects nearly similar features for predicting the ACE-FTS measurements. The most important feature across all altitude and latitude bands is consistently the corresponding tracer from the TOMCAT CTM. This finding strongly suggests that the fundamental controlling

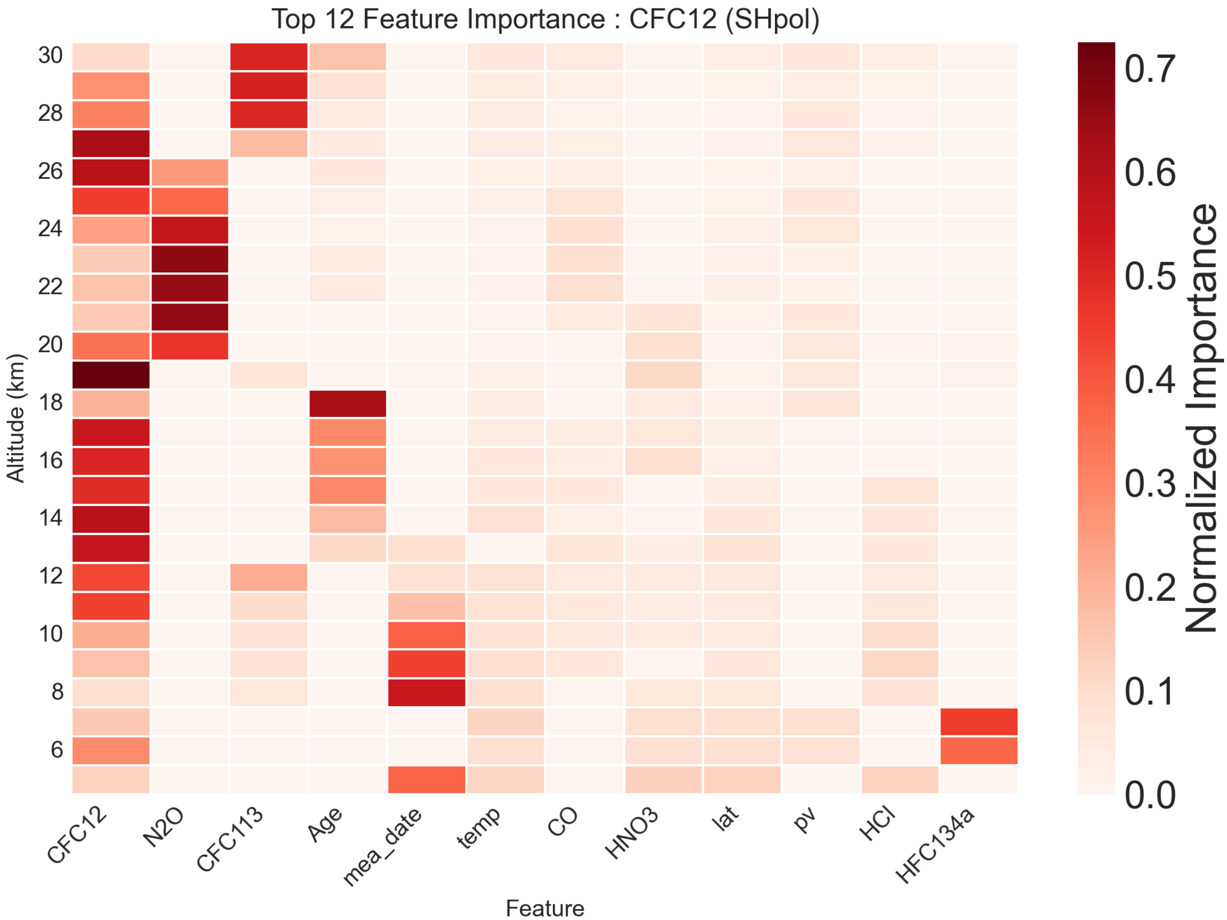

Figure 4. Same as Figure 3 but for CFC-12 and different y axis range, showing the altitude-dependent normalised relative importance of 12 most important features for the CFC-12 prediction.

processes for CFC ~~concentrations~~VMRs , such as transport and chemical loss, are well-represented within the TOMCAT CTM. Beyond the primary TOMCAT tracer feature, the remaining features show subtle, altitude-dependent differences:

– For CFC-11, the second and third most important features are the AoA (a proxy for transport) in the 14-18 km range, and the measurement date in the 6-9 km range, which reflects both dynamical and chemical controls on tropospheric ~~concentrations~~VMRs .

– For CFC-12, two transport tracers, N_2O and AoA, are selected within the 14-26 km altitude range, while the measurement date becomes more ~~significant~~ important below 10 km.

This general pattern remains consistent for low- to mid-latitude bands (e.g., SHmid, Trop, NHmid), where long-lived tracers (N_2O, $CFC-113$, HF, CFC-11, CFC-12) frequently appear as key secondary or tertiary features, particularly within the middle stratosphere. A ~~significant~~ notable exception is observed at higher latitudes (SHpol and NHpol), where the measurement date emerges as a somewhat important major feature, especially at lower (tropospheric) altitudes, compared to mid-low latitudes. The increased reliance on the measurement date at high latitudes suggests a deficiency in the TOMCAT simulation's ability to fully capture variations driven by crucial dynamical processes, especially stratosphere-troposphere exchange (STE). At high latitudes, the primary transport process is the strong seasonal downward transport, ~~specifically~~ resulting in the influx of CFC-poor air (due to stratospheric photolysis) from higher levels into the lower atmosphere, especially within the wintertime well-isolated polar vortex. This influx of older, CFC-depleted air likely requires the machine learning model to use the measurement date as an explicit correction term, indicating a need for TOMCAT to better resolve ~~high-latitude STE~~ downward descent from high altitudes as well as horizontal mixing.

5.2 Bias Correction Performance and SHAP Value Analysis

Scatterplots comparing the XGBoost-predicted (TCOM) and TOMCAT-simulated CFC-11 and CFC-12 ~~performance~~ performances with respect to ACE-FTS observations at 14 km for the SHpol latitude bin are shown in Figures 5 and 6, respectively. Similar analyses for SHmid (16 km), Trop (18 km), NHmid (16 km), and NHpol (14 km) are shown in Supplementary Figures S13 to S20. These figures clearly demonstrate the effectiveness of the developed bias-correction methodology. The corrected TCOM data points (orange circles) are visibly clustered much closer to the $1:1$ line compared to the TOMCAT data (blue circles), signifying a substantial reduction in systematic discrepancies. Pearson correlation coefficients, shown in the legend, also exhibit significant improvements. For example, in Figure 5 ~~$r = 0.70$~~ r increases from 0.70 for TOMCAT to $r = 0.94$ for TCOM, whereas in Figure 6 it improves from $r = 0.61$ to $r = 0.92$. The TCOM data points also include estimated uncertainties derived using quantile regression (the 25th and 75th percentiles).

~~Overall, the largest corrections seem to be applied at lower CFC VMRs, which most probably occurs during the late winter/early spring season. This is the period when strong downward transport of CFC-poor air occurs via the descending branch of the Brewer–Dobson circulation. In this regime, TOMCAT systematically shows much lower values than ACE-FTS suggests. The most probable reason for this model–measurement disagreement is that there are discrepancies in the TOMCAT~~

~~dynamical setup or biases in the stratospheric circulation in ERA5, failing to adequately supply the polar region with CFC-rich air.~~

Figures 5 and 6 also show the SHAP (SHapley Additive exPlanations) values for the XGBoost model at 14 km. SHAP analysis is a crucial component of interpretable machine learning, providing a local, additive explanation for each prediction. Unlike the overall feature importance metric, SHAP values quantify the contribution of each input feature to a single, specific prediction, including the direction (positive or negative) of the influence.

~~(a) Comparison of CFC-11 VMRs (in pptv) at 14 km altitude in the Southern Hemisphere polar (SHpol) latitude band. The scatter plot compares ACE-FTS observations (horizontal axis) against the model output: TOMCAT CTM (blue circles) and XGBoost regression model estimated VMRs (TCOM, orange circles). The vertical lines on the TCOM data represent the estimated uncertainties derived using quantile regression (25th and 75th percentiles). The dashed grey line indicates the 1:1 line to illustrate deviations with respect to ACE-FTS data. The legend also includes the Pearson correlation between ACE-FTS and the TOMCAT or TCOM data. (b) SHAP (SHapley Additive exPlanations) values for the XGBoost model at the same level (14 km). The SHAP values indicate the relative importance of each input feature and the direction (positive or negative) of their influence on the XGBoost model's prediction (corrected as TCOM).~~

~~Same as Figure 5, but for CFC-12.~~

~~For both CFC-11 and CFC-12 at 14 km in the SHpol region , the SHAP analysis reveals that the XGBoost bias correction model primarily acts~~ The analysis of the SHAP values in Figure 5b for CFC-11 at 14 km in the SHpol region provides a detailed look at how XGBoost functions as a transport ~~and tracer relationship corrector . For CFC-11, the most influential features are AoA, the TOMCAT CFC-11 tracer , and H_2O. High AoA (older air) and low TOMCAT CFC-11 concentrations (simulated as highly depleted) are both strongly associated with a positive SHAP value, which means the model is pushing the final predicted concentration higher. This suggests that~~ corrector by applying specific adjustments based on the chemical and dynamical characteristics of different air masses. The results indicate that XGBoost is able to identify distinct regimes to rectify biases in the underlying TOMCAT fields. Age-of-Air is identified as the most influential dynamical feature, where high values representing photochemically aged air that has descended from the mid-to-upper stratosphere are associated with negative SHAP values reaching -50 pptv. Conversely, younger air masses receive positive adjustments of up to $+40$ pptv. The simulated TOMCAT CFC-11 tracer concentration itself also plays a major role; air masses where TOMCAT-simulated VMRs are relatively high, typically representing air transported from lower latitudes, receive positive corrections of approximately $+40$ pptv, while very little adjustment occurs at lower concentration levels. Other features also help to characterise these regimes. For example, air masses with high water vapour and warmer temperatures receive negative adjustments, whereas corresponding air with low water vapour (a sign of dehydrated vortex air), and colder conditions receive positive adjustments. These patterns suggest that the largest downward adjustments occur when the air ~~mass bears the signature of being heavily processed and descended from the mid-upper stratosphere, TOMCAT typically underestimates the residual concentration relative to ACE – FTS measurements, requiring an upward correction~~is older and drier, indicating that TOMCAT tends to overestimate residual VMRs in heavily photochemically aged air descended during winter. This typically occurs near the polar vortex edge or after breakup where simulated vertical descent and horizontal mixing are not fully captured in the raw TOMCAT fields.

~~A similar pattern is observed for CFC-12, though the most influential features are ordered slightly differently: the TOMCAT CFC-12 tracer , AoA, and temp. The TOMCAT tracer concentration itself is the most important feature, reflecting the species' slower loss profile, which makes the absolute concentration a critical indicator of model performance. As with CFC-11, low TOMCAT CFC-12 concentrations (simulated as depleted) are associated with a positive SHAP value. This indicates that the XGBoost correction frequently adjusts the simulated concentrations upward when TOMCAT predicts greater photochemical destruction than observed by ACE-FTS for these highly processed air parcels.~~

~~The influences of the third most important variables further indicate unique characteristics of the polar vortexenvironment. For CFC-11, the importance of low H_2O values, which characterise the dehydrated air masses within the vortex, drives a positive correction. For CFC-12, the importance of~~ Similar distinct adjustment patterns are observed for CFC-12 as shown in Figure 6b, where the tracer concentration, Age-of-Air, and temperature are the most influential features. Air masses with a high Age-of-Air and low simulated TOMCAT VMRs are strongly associated with negative SHAP values, reaching approximately -50 and -40 pptv, respectively. This confirms a downward adjustment to the XGBoost predicted mixing ratios, suggesting that TOMCAT tends to overestimate residual CFC-12 in aged air that has descended within the polar vortex. In contrast, low ~~temperature values, which define the colder conditions of the~~ temperatures, which are a defining characteristic of the cold polar stratosphere, ~~similarly drives a positive correction. In both cases, the model correction is linked with the thermodynamic~~ drive large positive corrections of up to $+60$ pptv. This highlights how the model utilises thermodynamic signatures to rectify deficiencies in the simulated stratospheric circulation and its representation in meteorological reanalyses.

The scatterplots and SHAP analyses for the remaining latitude bands, provided in Supplementary Figures S13 to S20, confirm the global robustness and consistency of the TCOM framework. Across all latitude bins—SHmid (16 km), Trop (18 km), NHmid (16 km), and NHpol (14 km)—the corrected TCOM data points consistently exhibit a tighter clustering around the 1:1 line, with significant improvements in Pearson correlation coefficients relative to the baseline TOMCAT simulations. Consistent with the SHpol analysis, the SHAP results for these regions identify the corresponding TOMCAT tracer as a major predictor, though in the tropical upper troposphere (Fig. S15), other features such as CFC-113 and temperature exhibit higher relative importance. Additionally, the high relative influence of dynamical features such as Age-of-Air, temperature, and potential vorticity across all latitude bins underscores the model's role in rectifying circulation-driven biases. In the Trop region (Figs. S15 and ~~chemical environment of the vortex, reinforcing the conclusion that the XGBoost is refining the representation of CFC vertical descent and horizontal mixing that is inadequately captured by TOMCAT forced by ERA5 alone. Thus, the SHAP analysisat 14 km again suggest that the XGBoost primarily act as a transport corrector. By prioritising air mass age and characteristics of the TOMCAT tracer field, the model identifies and corrects for deficiencies in how TOMCAT simulates stratospheric circulation, descent, and mixing—particularly within the dynamically isolated polar vortex.~~ S16), where dynamical variability is large and observational constraints are relatively sparse, TCOM effectively resolves notable TOMCAT offsets, with correlation coefficients improving from 0.82 to 0.93 for CFC-11 and 0.76 to 0.89 for CFC-12. In the NHmid and NHpol regions (Figs. S17 to S20), the bias-correction patterns mirror those observed in the SH, reinforcing the conclusion that

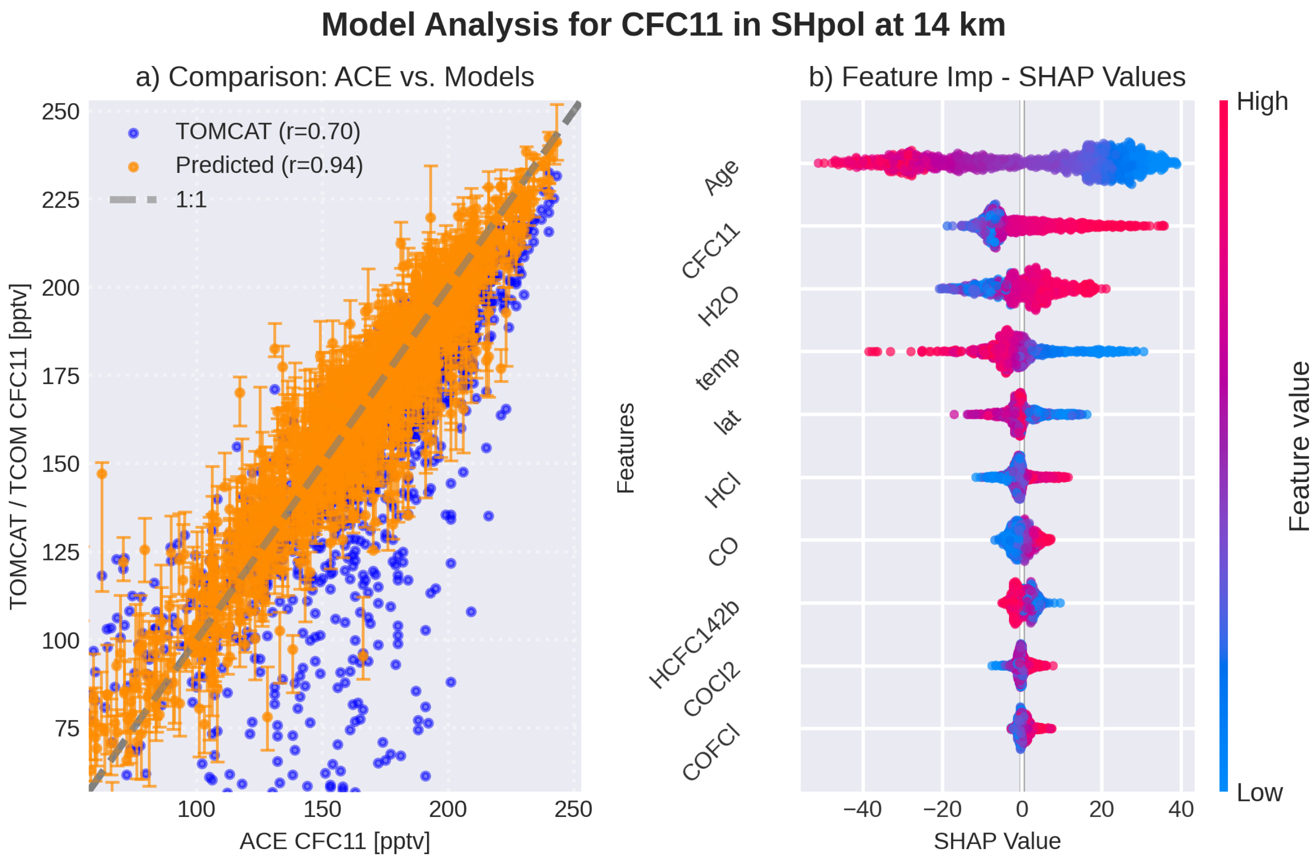

Figure 5. (a) Comparison of CFC-11 VMRs (in parts per trillion by volume or pptv) at 14 km altitude in the Southern Hemisphere polar (SHpol) latitude band. The scatter plot compares ACE-FTS observations (horizontal axis) against the model output: TOMCAT CTM (blue circles) and XGBoost regression model estimated VMRs (TCOM, orange circles). The vertical lines on the TCOM data represent the estimated uncertainties derived using quantile regression (25th and 75th percentiles). The dashed grey line indicates the 1:1 line to illustrate deviations with respect to ACE-FTS data. The legend also includes the Pearson correlation between ACE-FTS and the TOMCAT or TCOM data. (b) SHAP (SHapley Additive exPlanations) values for the XGBoost model at the same level (14 km). The SHAP values indicate the relative importance of each input feature and the direction (positive or negative) of their influence on the XGBoost model's prediction (corrected as TCOM).

the TCOM methodology successfully harmonises sparse satellite observations with continuous model background fields across range of atmospheric regimes.

5.3 Analysis of TCOM vs TOMCAT Differences

We now analyse percentage differences between TCOM and TOMCAT VMRs for CFC-11 and CFC-12. Figures 7 and 8 reveal distinct patterns in the corrections estimated by the XGBoost model across different altitudes, latitudes, and time. A common

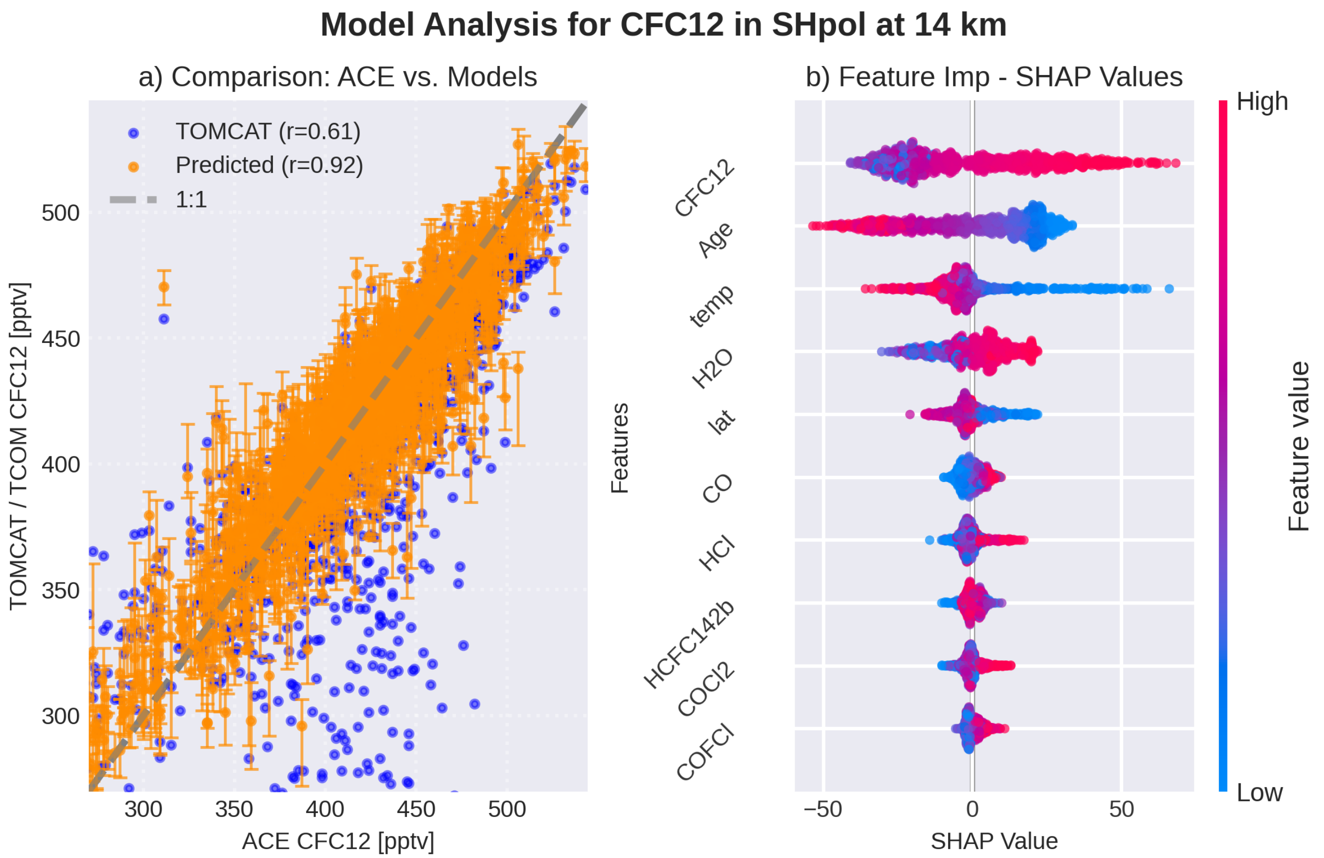

Figure 6. Same as Figure 5, but for CFC-12.

feature for both CFC-11 and CFC-12 is the presence of large positive and negative percentage differences, particularly in the high-latitude regions (south of ~~50°S~~ 60°S and north of ~~50°N). At~~ 60°N); where XGBoost seems to adjust TCOM profiles upwards. At mid-latitude (around 40°), XGBoost seems to remove positive biases in TOMCAT, whereas near tropical latitudes negative biases are corrected. At higher altitudes (25 km and 20 km), large differences are also observed in the tropics. This indicates that the XGBoost correction, as captured in the TCOM dataset, introduces significant adjustments compared to the TOMCAT baseline, especially in areas characterised by polar vortex dynamics and seasonal transport coinciding with extreme polar events, such as the anomalous Antarctic winter of 2002 (e.g. Weber et al., 2003) or the severe Arctic winters in 2020 (e.g. Feng et al., 2021) and 2024 (e.g. Newman et al., 2024). The extreme correction, sometimes exceeding ±80% or even ±100% in the 25 km panel for CFC-12, underscores the magnitude of the correction needed in these chemically and dynamically active upper stratospheric levels.

At lower altitudes, specifically 15 km and 10 km, the scale of the percentage differences for both CFC-11 and CFC-12 decreases substantially. Maximum differences are generally constrained to ±40% and ±20% at 15 km and 10 km, respec-

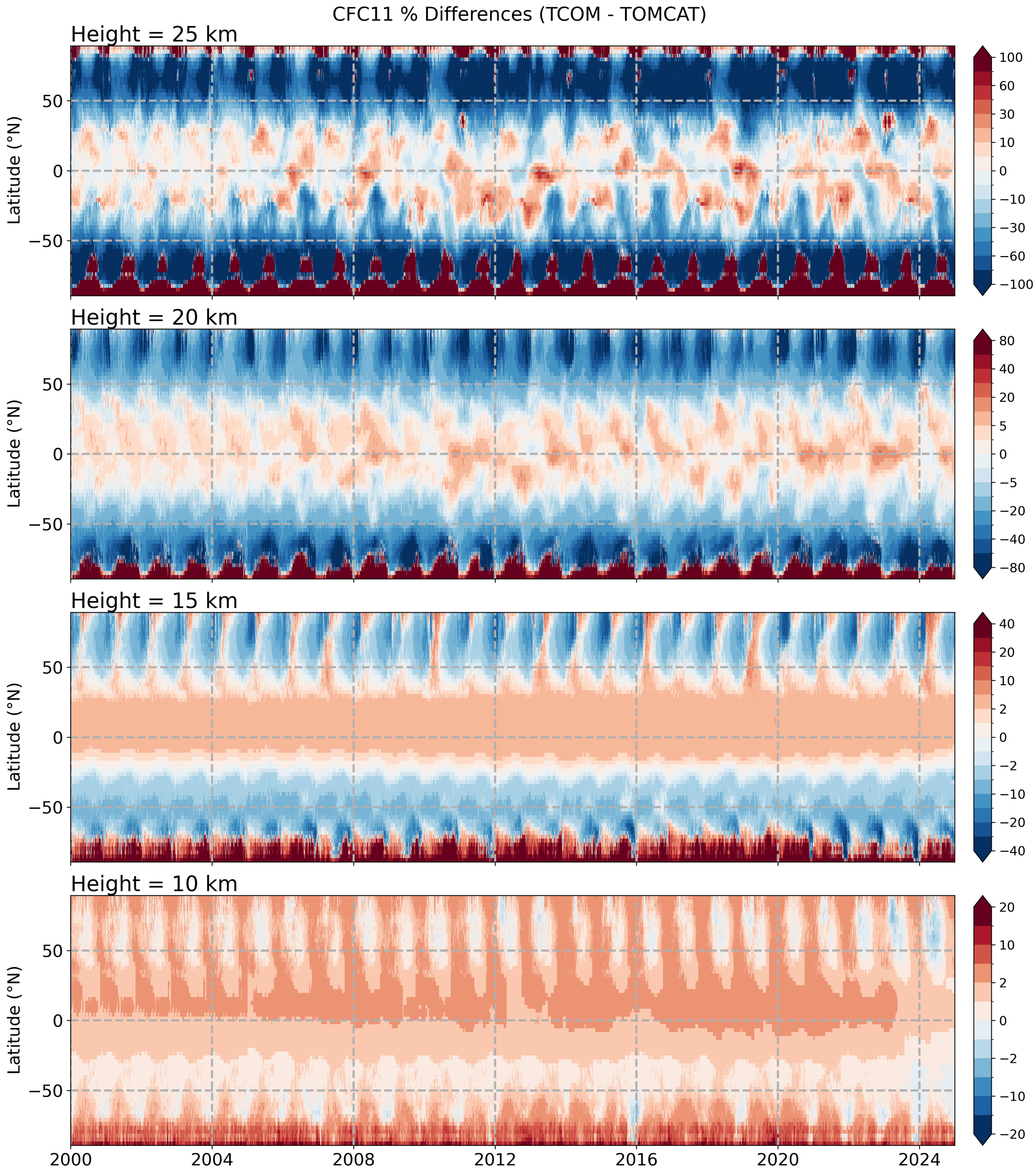

Figure 7. Percentage differences between the TCOM and TOMCAT VMRs of CFC-11 as a function of time (2000–2024) and latitude (85°S to 85°N). Differences are presented for four altitude levels: 25 km (top panel), 20 km, 15 km, and 10 km (bottom panel). The differences are calculated as $100 \times (\mathrm{TCOM} - \mathrm{TOMCAT})/\mathrm{TCOM}$. The ~~colour bar scale~~ colour-bar scales for the percentage difference ~~is~~ are irregular and optimised for each individual altitude panel.

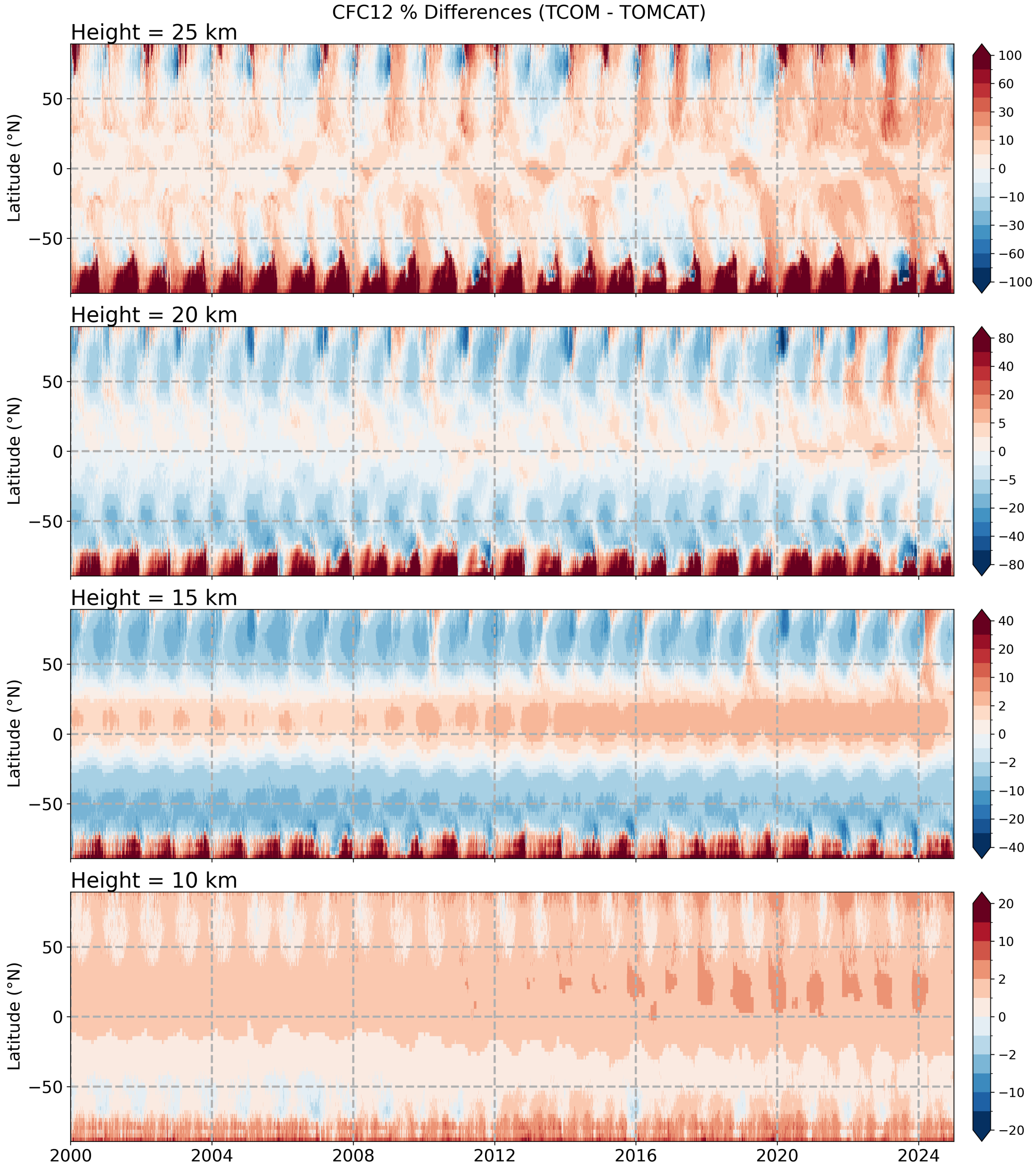

Figure 8. Same as Figure 7, but for CFC-12, comparing TOMCAT and TCOM data.

tively. This suggests that TOMCAT's performance, when compared to the TCOM product, is more consistent in the lower stratosphere, implying that the fundamental chemical and transport processes for these CFCs are reasonably well simulated by the CTM in this region. However, distinct geographical and temporal patterns persist. At $10\ \mathrm{km}$, a band of sustained positive difference is visible across the tropical and mid-latitude regions ($50°\mathrm{S}$ to $50°\mathrm{N}$) throughout the entire 2000–2024 period for both CFCs. This implies a consistent, relatively small but persistent underestimation of CFC VMRs by TOMCAT in the tropical lower stratosphere, which XGBoost effectively rectifies. This persistent difference may stem from TOMCAT's photolysis loss occurring at lower altitudes than observations suggest. An interesting feature at lower altitudes is the inter-hemispheric asymmetry at low-mid latitudes. Persistent positive bands of difference are observed from the tropics to the NH mid-latitudes, particularly between 10 and 15 km (see Figs. 7 and 8). This appears to be consistent with the results of Schmidt et al. (2024), who analysed ACE-FTS data sets for various long-lived species and identified a distinct hemispheric asymmetry. They found that SH VMRs in the lower stratosphere exhibit a 2.25-year lag compared to those in the NH. They noted similar lags in NOAA surface observations and suggested that these differences may be linked to the concentration of industrial emissions in the NH, the time required for these changes to mix thoroughly into the measured atmospheric regions, or a combination of complex vertical and horizontal transport pathways. Furthermore, the strong gradient in differences near the edge of the polar vortex (especially in the SHpol region) ~~confirms~~ indicates that horizontal transport is likely too constrained in the TOMCAT simulation across almost all levels.

Additionally, there are noticeable differences in the correction pattern between the two compounds. For CFC-12, the regions of very large percentage differences at $25\ \mathrm{km}$ appear more concentrated and tightly coupled with the seasonal cycle in the polar regions, suggesting a stronger influence of the polar vortex and associated transport processes. In contrast, the CFC-11 plots show a broader, albeit still high-magnitude, correction pattern at $25\ \mathrm{km}$, with notable positive differences extending into the mid-latitudes across the entire time series. This may reflect differences in the photochemical lifetimes (e.g. Chipperfield et al., 2014; Hoffmann et al., 2014) and vertical distribution of the two CFCs, leading to distinct sensitivities in the XGBoost corrections.

On a long-term basis (2000 to 2024), there is minor but systematic divergence in the correction patterns for both species. For the top three levels, differences seem to grow larger with time, with the largest corrections estimated for recent years, which coincide with record minimum or maximum springtime ozone observations. The exact causes of these diverging patterns are not fully understood. However, unusual changes in stratospheric composition (e.g., the Australian New Year (ANY) fires in 2020 or the Hunga eruption in 2022) might have caused ~~unusual~~ changes in stratospheric circulation that were not accurately captured by the ERA5 reanalysis data used to drive TOMCAT. We aim to analyse these patterns in future studies.

5.4 Comparison with SPARC Data

~~The comparisons between TOMCAT, TCOM (model-corrected) dataset, and the~~ A critical evaluation of the TCOM dataset is provided through comparisons with independent observation-based ~~data from the SPARC~~ products from the SPARC v2 compilation (Hegglin et al., 2021)~~for three distinct pressure levels are shown in Figures 9 and 10. As previously noted (e.g. Section 5.2)~~, and is presented in Figure 9 for CFC-11 and Figure 10 for CFC-12. To ensure clarity in the figures and to

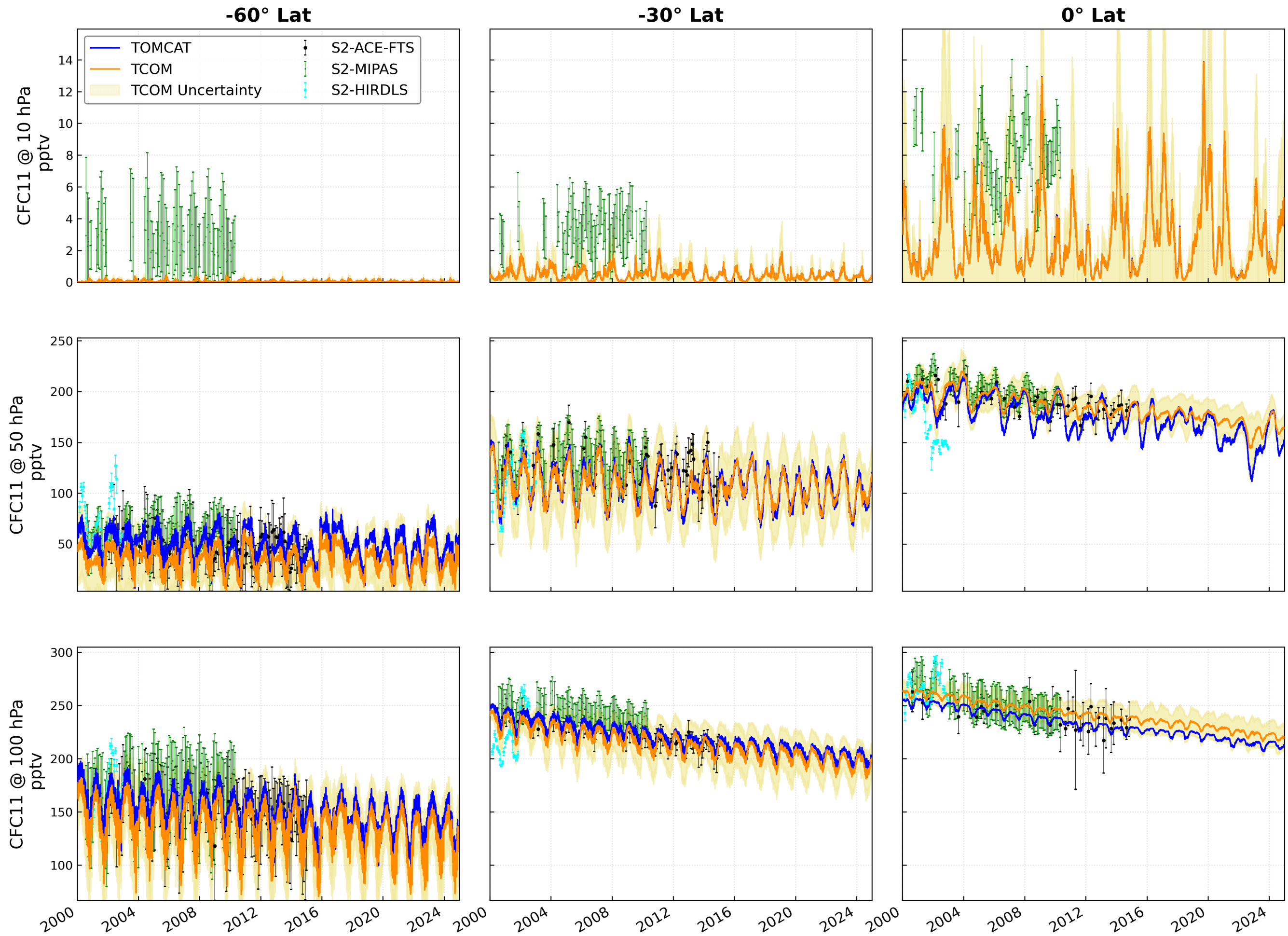

Figure 9. Temporal comparison of daily-mean CFC-11 VMRs (in pptv) from 2000 to 2024 across ~~five~~ three distinct latitude bands (~~e.g.~~60°S, ~~−67.5° to 67.5°~~30°S and 0°) at three stratospheric pressure levels: 10 hPa (≈ 30 km), 50 hPa (≈ 20 km), and 100 hPa (≈ 16 km). Data sources shown include: the TOMCAT CTM output (solid blue lines), the XGBoost-predicted TCOM dataset (solid orange lines), and TCOM uncertainty (±1σ derived from quantile regression, light orange shading). Monthly mean data points from observation-based data from the SPARC v2 (Hegglin et al., 2021) are also shown. Data points and related standard deviations from ACE-FTS, MIPAS and HIRDLS are shown with black, green and aqua coloured dots, respectively.

analyse different dynamical regimes in detail, we focus on the three representative latitude bins: high-latitude (SHpol, 60°S), subtropical (30°S), and tropical (0°), across three stratospheric pressure levels (100, ~~these figures clearly show~~ 50, and 10 hPa).

In the middle stratosphere at 50 hPa, a systematic low bias in ~~TOMCAT in the middle stratosphere, particularly evident for the middle three panels (32.5° S, 2.5° N, and 32.5° N) at the 50 hPa pressure level~~raw TOMCAT output is most evident,

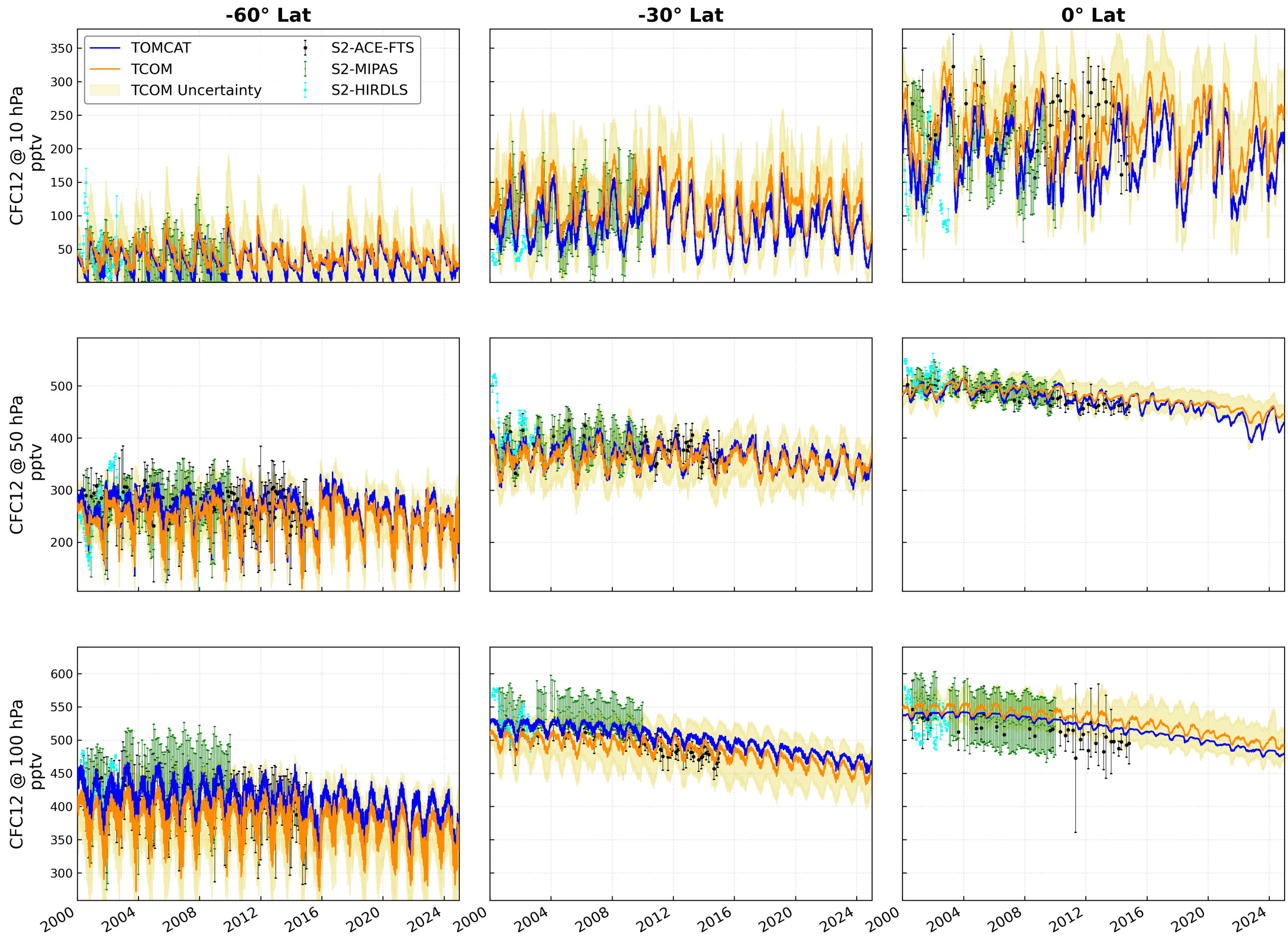

Figure 10. Same as Figure 9, but for CFC-12.

particularly in the tropical and subtropical panels. This consistent offset suggests that ~~TOMCAT is misrepresenting the~~ while TOMCAT accurately represents the large-scale declining trend, it misrepresents the vertical gradient of these long-lived species~~in the lower-to-middle stratosphere. Specifically, while the 100 hPa panels show good agreement, indicating that the model 's STE is broadly correct, the lower observed VMRs at 50 hPa imply that the model either transports CFC-rich air too quickly upwards or fails to efficiently mix this air horizontally .~~, possibly due to model deficiencies in representing the horizontal mixing of CFC-rich air from the tropical pipe into ~~the~~ mid-latitudes~~, leading to a noticeable deficit in these regions. The XGBoost correction , which forms the TCOM dataset, .~~ The TCOM correction successfully resolves this bias~~by modifying the VMRs and aligning them with the observations~~, aligning the mixing ratios with the satellite-observations-based SPARC v2 monthly mean data points.

~~The analysis also reveals significant differences between various satellite measurement-based data products, and the consistencies/inconsi across these datasets highlights the relevance of TCOM for both model evaluation and understanding~~

5.4.1 CFC-11 Comparison Analysis

The CFC-11 comparison in Figure 9 highlights the efficacy of the machine learning approach across varying dynamical regimes:

- **Tropics (0°)**: At 100 hPa and 50 hPa, TOMCAT exhibits a significant negative bias relative to the high-density cluster of ACE-FTS, MIPAS, and HIRDLS measurements. TCOM effectively bridges this gap, while at the highly variable 10 hPa level, it tracks the sparse satellite scatter and captures the long-term ~~changes in the abundance of these CFCs. In the lower stratosphere (100 hPa), all the datasets (TOMCAT, TCOM, ACE-FTS, MIPAS, and HIRDLS) show the highest degree of agreement in absolute VMRs and the declining long-term trend. This high consistency reflects the effectiveness of the Montreal Protocol, and the lower concentrations at high latitudes show the dominant role of the large-scale transport at this level. Overall, observational datasets generally overlap closely, particularly in the tropical and mid-latitude bands (32.5° S to 32.5° N).~~

  ~~At 50 hPa, discrepancies between the model and observations become more pronounced. However, the satellite datasets remain largely consistent with each other in these regions,~~ decline indicated by MIPAS.

- **Subtropics (30°S)**: The 100 hPa panel shows very good agreement across all datasets, confirming that model-simulated stratosphere-troposphere exchange (STE) is broadly correct in this region. At 100 hPa and 50 hPa, TCOM applies minor downward and upward adjustments respectively, and ~~TCOM consistently remains within their uncertainty estimates. For example, at 32.5° N for both CFC-11 and CFC-12, ACE-FTS (black) and MIPAS (green/aqua) track each other closely, confirming the magnitude of the bias that TCOM successfully removes.~~

  ~~At~~ it matches the key features in SPARC ACE-FTS data points, and also identifies a persistent high bias in the MIPAS data points relative to the ACE-FTS.

- **Mid-high latitudes (60°S)**: Strong seasonal oscillations are visible at 100 hPa, where TCOM adjusts downwards to match observational constraints. At the ~~10 hPa level, the largest influence of natural variability coupled with the highest uncertainty is visible across all datasets. This is consistent with the highly dynamic and chemically active nature of these species in this region. While the TCOM correction introduces substantial short-term variability, aiming to capture the~~ 50 hPa level, the XGBoost correction is vital for rectifying the overestimation of TOMCAT values during winter correcting either weaker polar descent or horizontal mixing near the edge of the vortex, ensuring the final product accurately reflects the chemically processed air within the high-latitude vortex.

5.4.2 CFC-12 Comparison Analysis

Similar robust improvements are observed for CFC-12 in Figure 10, which benefits from better retrievals at higher altitudes:

– **Tropics (0°)**: The 100 hPa tropical panel shows the most prominent deficiencies in TOMCAT data, which is systematically removed in TCOM to match the SPARC data points. At 50 hPa, the adjustment is minor. At 10 hPa, where natural variability is highest, TCOM introduces daily fluctuations that successfully mirrors the dynamic ~~short-term cycles seen in the observations, the ACE-FTS~~ cycle captured by MIPAS and ACE-FTS.

– **Subtropics (30°S)**: At 100 hPa all the data sets confirm declining trends and TCOM shows downward adjustment to match SPARC data points. The 50 hPa level reveals that TCOM remains centred within the ACE-FTS scatter (black dots)~~and MIPAS/HIRDLS data (green/aqua dots) themselves show the large scatter and relative differences compared to lower altitudes. For CFC-11 (Figure 9), where concentrations are nearly negligible at 10 hPa, the observational data is sparser and more variable, particularly in the polar regions (67.5° S and 67.5° N), but TCOM successfully tracks the general long-term decline indicated by the MIPAS data (green dots).~~ , consistently staying within uncertainty estimates while TOMCAT data suggests very little adjustment is needed.

– **Mid-high latitudes (60°S)**: The polar panels also show that TCOM data points remain close to the SPARC ACE-FTS data points. At 100 hPa and 50 hPa, the downward correction accounts for the TOMCAT's inability to fully simulate the descent of older air (or horizontal mixing), bringing TCOM into better alignment with both MIPAS and ACE-FTS monthly means. On the other hand, at 10 hPa upward adjustment is needed to match SPARC data points.

Overall, this inter-comparison confirms that ~~the~~ while different satellite measurements ~~of CFC profiles are highly consistent , especially in the middle and lower stratosphere. The primary value of the TCOM dataset is its ability to statistically harmonise the TOMCAT~~ are generally consistent in the lower stratosphere, the TCOM dataset provides unique added value by statistically harmonising continuous model output with these rigorous observational constraints, ~~thus providing a consistent , gap-free, and corrected dataset.~~ resulting in a consistent and high-fidelity global record

6 Summary and Conclusions

We present the TCOM-CFC-11 and TCOM-CFC-12 (~~version 2.0~~v2.0) datasets, which provide continuous, gap-free, global daily vertical profiles of CFC VMRs from 2000 to 2024. This was achieved using an innovative methodology that employs XGBoost regression to constrain the continuous output of the TOMCAT CTM against high-quality ACE-FTS satellite observations, thereby addressing the limitations of sparse satellite coverage. The comparative analysis demonstrated the method's effectiveness, as the corrected TCOM data clustered significantly closer to the observations, effectively removing the systematic low bias present in TOMCAT in the middle stratosphere (e.g., at 50 hPa) when validated against independent observation-based products such as the SPARC v2 data set.

The analysis of model diagnostics, particularly the SHAP ~~(SHapley Additive exPlanations)~~ values, provided crucial insight into the physical origin of the deficiencies of the TOMCAT simulation. The feature importance investigation strongly suggests that the XGBoost regression model functions primarily as a "transport corrector". This is evidenced by the high influence of dynamical features, such as ~~Age~~age-of-~~Air~~ air (AoA) and ~~Potential Vorticity~~ potential vorticity (PV), indicating that the main source of model bias relates to how TOMCAT simulates stratospheric circulation, rather than its chemical schemes. This transport correction is most notable in the polar regions, where TCOM rectifies a systematic underestimation in TOMCAT~~, Most probable reason are~~ . The most probable reason is likely to be errors in ~~transport related~~ transport-related parameters in TOMCAT or biases in the representation of the stratospheric circulation in the ERA5 reanalysis fields that are used to drive the model.

In conclusion, the TCOM datasets successfully harmonise the TOMCAT CTM output with rigorous observational constraints, yielding a consistent, high-fidelity resource for the stratospheric research community. Version 2.0 incorporates the latest ACE-FTS v5.3 data and provides robust uncertainty estimates derived from ensemble retraining using the 25th and 75th percentiles of observations. The final TCOM-CFC-11 and TCOM-CFC-12 datasets ~~will be publicly released in December 2025. They~~ are (since January 2026) publicly released and will serve as a valuable, observationally-constrained benchmark for evaluating and refining CTMs and CCMs, helping to reduce uncertainties in long-term trend analyses of ozone-depleting substances, and aiding in the assessment of the Montreal Protocol's effectiveness.

Finally, while the current version of the TCOM dataset relies on the high-quality constraints provided by ACE-FTS, the methodology is designed to be adaptable to future observational streams. The upcoming ALICE instrument on NASA STRIVE mission (to be launched in early 2030s), with its ability to measure vertical profiles for a wide range of trace gases, including both CFCs and other species currently included in the TCOM datasets (e.g. O_3, CH_4, N_2O, HNO_3, H_2O), will offer a critical opportunity to maintain and extend these gap-free data records well into the next decade. Within the current framework, we have shown that even with a data gap (e.g. 2000 to 2003 in this study) profile data sets can be constructed for most of the long-lived species based on the chemical-dynamical information derived during the observed period. Extrapolations based on gradually varying source gas loadings, e.g. obtainable from ground-based observations, can be accommodated. The most problematic issue would be sudden atmospheric perturbations such as the changes in the stratospheric aerosol properties due to volcanic eruptions or wildfires for which direct aerosol observations are needed. Nonetheless, future trace observations from missions such as STRIVE are essential to provide reference anchor points for the end of a gap-filled period.

7 Data availability

The TCOM version 2.0 datasets are publicly available on Zenodo. While the initial methodology for version 1.0 was detailed in Dhomse and Chipperfield (2023) focusing on CH_4 and N_2O, the current v2.0 products described in this study have been expanded to include a range of stratospheric species. Specifically, the datasets for CFC-11 and CFC-12 can be accessed at https://zenodo.org/records/18145730 (Dhomse, 2026a) and https://zenodo.org/records/18147392 (Dhomse, 2026b), respectively. Using the same methodology, v2.0 profile datasets are also publicly available for O_3 (https://zenodo.org/records/18199586), H_2O (https://zenodo.org/records/18199962), CH_4 (https://zenodo.org/records/18197333), N_2O (https://zenodo.org/records/18197444), HF (https://zenodo.org/records/18184779), HCl (https://zenodo.org/records/18184430), HNO_3 (https://zenodo.org/records/18199002), and COF_2 (https://zenodo.org/records/18201786). These links are all listed and maintained on the overview TCOM page (https://tomcat.leeds.ac.uk/tomcat/tcom).

The ACE-FTS v5.3 data utilised for these TCOM constraints are available at the official archive: https://databace.scisat.ca. Additional daily 3D profile fields, for both TOMCAT and TCOM are available from the corresponding author upon request.

Author contributions. SSD conceived the idea, performed the analysis, and wrote the draft paper. MPC performed TOMCAT model simulations, provided the model output, and contributed to writing the paper. This project builds upon the work initiated by Dhomse et al. (2021) and Dhomse and Chipperfield (2023).

Competing interests. The authors declare that they have no competing interests.

Disclaimer. TEXT

Acknowledgements. SSD and MPC were supported by the Natural Environment Research Council (NERC) LSO3 (NE/V011863/1) and InHALE (NE/X003450/1), and NCEO TerraFIRMA projects. The Atmospheric Chemistry Experiment (ACE), also known as SCISAT, is a Canadian-led mission mainly supported by the Canadian Space Agency. We thank the European Centre for Medium-Range Weather Forecasts for providing their analyses. TOMCAT simulations were performed on the UK national Archer2 and Leeds Arc4/Aire HPC systems.

References

Bernath, P.: Atmospheric Chemistry Experiment (ACE): An overview, IEEE International Geoscience and Remote Sensing Symposium, 2, 147–160, http://eprints.whiterose.ac.uk/68898/, 2002.

Bernath, P. F., McElroy, C. T., Abrams, M., Boone, C. D., Butler, M., Camy-Peyret, C., Carleer, M., Clerbaux, C., Coheur, P.-F., Colin, R., et al.: Atmospheric chemistry experiment (ACE): mission overview, Geophysical research letters, 32, 2005.

Boone, C., Bernath, P., Cok, D., Jones, S., and Steffen, J.: Version 4 retrievals for the atmospheric chemistry experiment Fourier transform spectrometer (ACE-FTS) and imagers, Journal of Quantitative Spectroscopy and Radiative Transfer, 247, 106 939, 2020.

Boone, C., Bernath, P., and Lecours, M.: Version 5 retrievals for ACE-FTS and ACE-imagers, Journal of Quantitative Spectroscopy and Radiative Transfer, 310, 108 749, 2023.

Boone, C. D., Nassar, R., Walker, K. A., Rochon, Y., McLeod, S. D., Rinsland, C. P., and Bernath, P. F.: Retrievals for the atmospheric chemistry experiment Fourier-transform spectrometer, Applied optics, 44, 7218–7231, 2005.

Boone, C. D., Walker, K. A., and Bernath, P. F.: Version 3 retrievals for the atmospheric chemistry experiment Fourier transform spectrometer (ACE-FTS), The atmospheric chemistry experiment ACE, 10, 103–127, 2013.

Bourguet, S., Stone, K., and Lickley, M.: Semi-empirical estimates of stratospheric circulation and the lifetimes of chlorofluorocarbons and carbon tetrachloride, Communications Earth & Environment, 6, 531, 2025.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Brown, A. T., Volk, C. M., Schoeberl, M. R., Boone, C. D., and Bernath, P. F.: Stratospheric lifetimes of CFC-12, CCl_4, CH_4, CH_3Cl and N_2O from measurements made by the Atmospheric Chemistry Experiment-Fourier Transform Spectrometer (ACE-FTS), Atmospheric Chemistry and Physics, 13, 6921–6950, https://doi.org/10.5194/acp-13-6921-2013, 2013.

Butchart, N. and Scaife, A. A.: Removal of chlorofluorocarbons by increased mass exchange between the stratosphere and troposphere in a changing climate, Nature, 410, 799–802, https://doi.org/10.1038/35071047, 2001.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, pp. 785–794, ACM, https://doi.org/10.1145/2939672.2939785, 2016.

Chipperfield, M. P.: Multiannual simulations with a three-dimensional chemical transport model, Journal of Geophysical Research, 104, 1781, https://doi.org/10.1029/98JD02597, 1999.

Chipperfield, M. P.: New version of the TOMCAT/SLIMCAT off-line chemical transport model: Intercomparison of stratospheric tracer experiments, Quarterly Journal of the Royal Meteorological Society, 132, 1179–1203, https://doi.org/10.1256/qj.05.51, 2006.

Chipperfield, M. P. and Bekki, S.: Opinion: Stratospheric ozone–Depletion, recovery and new challenges, Atmospheric Chemistry and Physics, 24, 2783–2802, 2024.

Chipperfield, M. P., Liang, Q., Strahan, S. E., Morgenstern, O., Dhomse, S. S., Abraham, N. L., Archibald, A. T., Bekki, S., Braesicke, P., Di Genova, G., et al.: Multimodel estimates of atmospheric lifetimes of long-lived ozone-depleting substances: Present and future, Journal of Geophysical Research: Atmospheres, 119, 2555–2573, 2014.

Chipperfield, M. P., Dhomse, S. S., Feng, W., McKenzie, R. L., Velders, G. J. M., and Pyle, J. A.: Quantifying the ozone and ultraviolet benefits already achieved by the Montreal Protocol., Nature communications, 6, 7233, https://doi.org/10.1038/ncomms8233, 2015.

Coddington, O., Lean, J. L., Pilewskie, P., Snow, M., and Lindholm, D.: A Solar Irradiance Climate Data Record, Bulletin of the American Meteorological Society, 97, 1265–1282, https://doi.org/10.1175/BAMS-D-14-00265.1, 2016.

Dhomse, S.: TCOM-CFC11 : TOMCAT CTM and Occultation Measurement based CFC-11 profile data set, https://doi.org/10.5281/zenodo.18145730, 2026a.

Dhomse, S.: TCOM-CFC12 : TOMCAT CTM and Occultation Measurement based CFC-12 profile data set, https://doi.org/10.5281/zenodo.18147392, 2026b.

Dhomse, S., Chipperfield, M., Damadeo, R., Zawodny, J., Ball, W., Feng, W., Hossaini, R., Mann, G., and Haigh, J.: On the ambiguous nature of the 11 year solar cycle signal in upper stratospheric ozone, Geophysical Research Letters, 43, 7241–7249, https://doi.org/10.1002/2016GL069958, 2016.

Dhomse, S., Feng, W., Montzka, S., Hossaini, R., Keeble, J., Pyle, J., Daniel, J., and Chipperfield, M.: Delay in recovery of the Antarctic ozone hole from unexpected CFC-11 emissions, Nature communications, 10, 5781, 2019.

Dhomse, S. S. and Chipperfield, M. P.: Using machine learning to construct TOMCAT model and occultation measurement-based stratospheric methane (TCOM-CH4) and nitrous oxide (TCOM-N2O) profile data sets, Earth System Science Data, 15, 5105–5120, 2023.

Dhomse, S. S., Chipperfield, M. P., Feng, W., Hossaini, R., Mann, G. W., and Santee, M. L.: Revisiting the hemispheric asymmetry in midlatitude ozone changes following the Mount Pinatubo eruption: A 3-D model study, Geophysical Research Letters, 42, 3038–3047, https://doi.org/10.1002/2015GL063052, 2015.

Dhomse, S. S., Kinnison, D., Chipperfield, M. P., Salawitch, R. J., Cionni, I., Hegglin, M. I., Abraham, N. L., Akiyoshi, H., Archibald, A. T., Bednarz, E. M., Bekki, S., Braesicke, P., Butchart, N., Dameris, M., Deushi, M., Frith, S., Hardiman, S. C., Hassler, B., Horowitz, L. W., Hu, R.-M., Jöckel, P., Josse, B., Kirner, O., Kremser, S., Langematz, U., Lewis, J., Marchand, M., Lin, M., Mancini, E., Marécal, V., Michou, M., Morgenstern, O., OConnor, F. M., Oman, L., Pitari, G., Plummer, D. A., Pyle, J. A., Revell, L. E., Rozanov, E., Schofield, R., Stenke, A., Stone, K., Sudo, K., Tilmes, S., Visioni, D., Yamashita, Y., and Zeng, G.: Estimates of ozone return dates from Chemistry-Climate Model Initiative simulations, Atmospheric Chemistry and Physics, 18, 8409–8438, https://doi.org/10.5194/acp-18-8409-2018, 2018.

Dhomse, S. S., Arosio, C., Feng, W., Rozanov, A., Weber, M., and Chipperfield, M. P.: ML-TOMCAT: machine-learning-based satellite-corrected global stratospheric ozone profile data set from a chemical transport model, Earth System Science Data, 13, 5711–5729, 2021.

Dhomse, S. S., Chipperfield, M. P., Feng, W., Hossaini, R., Mann, G. W., Santee, M. L., and Weber, M.: A single-peak-structured solar cycle signal in stratospheric ozone based on Microwave Limb Sounder observations and model simulations, Atmospheric Chemistry and Physics, 22, 903–916, 2022.

Eckert, E., Laeng, A., Lossow, S., Kellmann, S., Stiller, G., Von Clarmann, T., Glatthor, N., Höpfner, M., Kiefer, M., Oelhaf, H., et al.: MIPAS IMK/IAA CFC-11 (CCl 3 F) and CFC-12 (CCl 2 F 2) measurements: accuracy, precision and long-term stability, Atmospheric Measurement Techniques, 9, 3355–3389, 2016.

Eyring, V., Waugh, D. W., Bodeker, G. E., Cordero, E., Akiyoshi, H., Austin, J., Beagley, S. R., Boville, B. A., Braesicke, P., Brühl, C., Butchart, N., Chipperfield, M. P., Dameris, M., Deckert, R., Deushi, M., Frith, S. M., Garcia, R. R., Gettelman, A., Giorgetta, M. A., Kinnison, D. E., Mancini, E., Manzini, E., Marsh, D. R., Matthes, S., Nagashima, T., Newman, P. A., Nielsen, J. E., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Schraner, M., Scinocca, J. F., Semeniuk, K., Shepherd, T. G., Shibata, K., Steil, B., Stolarski, R. S., Tian, W., and Yoshiki, M.: Multimodel projections of stratospheric ozone in the 21st century, Journal of Geophysical Research, 112, D16 303, https://doi.org/10.1029/2006JD008332, 2007.

Eyring, V., Cionni, I., Bodeker, G. E., J., D. E. C.-P. A., Kinnison, Scinocca, J. F., Waugh, D. W., Akiyoshi, H., Bekki, S., Chipperfield, M. P., Dameris, M., Dhomse, S., Frith, S., Garny, H., Gettelman, A., Kubin, A., Langematz, U., Mancini, E., Marchand, M., Nakamura, T., Oman, L. D., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Shepherd, T. G., Shibata, K., Tian, W., Braesicke, P., Hardiman, S. C., Lamarque, J. F., Morgenstern, O., Pyle, J. A., Smale, D., and Yamashita, Y.: Multi-model assessment of stratospheric ozone return dates and ozone recovery in CCMVal-2 models, ATMOSPHERIC CHEMISTRY AND PHYSICS, 10, 9451–9472, https://doi.org/10.5194/acp-10-9451-2010, 2010.

Feng, W., Dhomse, S. S., Arosio, C., Weber, M., Burrows, J. P., Santee, M. L., and Chipperfield, M. P.: Arctic Ozone Depletion in 2019/20: Roles of Chemistry, Dynamics and the Montreal Protocol, GEOPHYSICAL RESEARCH LETTERS, 48, https://doi.org/10.1029/2020GL091911, 2021.

Fischer, H., Birk, M., Blom, C., Carli, B., Carlotti, M., von Clarmann, T., Delbouille, L., Dudhia, A., Ehhalt, D., Endemann, M., Flaud, J. M., Gessner, R., Kleinert, A., Koopman, R., Langen, J., López-Puertas, M., Mosner, P., Nett, H., Oelhaf, H., Perron, G., Remedios, J., Ridolfi, M., Stiller, G., and Zander, R.: MIPAS: an instrument for atmospheric and climate research, Atmospheric Chemistry and Physics, 8, 2151–2188, https://doi.org/10.5194/acp-8-2151-2008, 2008.

Freund, Y. and Schapire, R. E.: A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, Journal of Computer and System Sciences, 55, 119–139, https://doi.org/10.1006/jcss.1997.1504, 1997.

Gille, J., Barnett, J., Arter, P., Barker, M., Bernath, P., Boone, C., Cavanaugh, C., Chow, J., Coffey, M., Craft, J., et al.: High Resolution Dynamics Limb Sounder: Experiment overview, recovery, and validation of initial temperature data, Journal of Geophysical Research: Atmospheres, 113, 2008.

Hegglin, M. I., Tegtmeier, S., Anderson, J., Bourassa, A. E., Brohede, S., Degenstein, D., Froidevaux, L., Funke, B., Gille, J., Kasai, Y., Kyrölä, E. T., Lumpe, J., Murtagh, D., Neu, J. L., Pérot, K., Remsberg, E. E., Rozanov, A., Toohey, M., Urban, J., von Clarmann, T., Walker, K. A., Wang, H.-J., Arosio, C., Damadeo, R., Fuller, R. A., Lingenfelser, G., McLinden, C., Pendlebury, D., Roth, C., Ryan, N. J., Sioris, C., Smith, L., and Weigel, K.: Overview and update of the SPARC Data Initiative: comparison of stratospheric composition measurements from satellite limb sounders, Earth System Science Data, 13, 1855–1903, https://doi.org/10.5194/essd-13-1855-2021, 2021.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Hoerl, A. E. and Kennard, R. W.: Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics, 12, 55–67, https://doi.org/10.1080/00401706.1970.10488634, 1970.

Hoffmann, L., Hoppe, C. M., Müller, R., Dutton, G. S., Gille, J. C., Griessbach, S., Jones, A., Meyer, C. I., Spang, R., Volk, C. M., and Walker, K. A.: Stratospheric lifetime ratio of CFC-11 and CFC-12 from satellite and model climatologies, Atmospheric Chemistry and Physics, 14, 12 479–12 497, https://doi.org/10.5194/acp-14-12479-2014, 2014.

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., and Pfister, L.: Stratosphere-troposphere exchange, Reviews of Geophysics, 33, 403–439, https://doi.org/10.1029/95RG02097, 1995.

Khosravi, R., Lambert, A., Lee, H., Gille, J., Barnett, J., Francis, G., Edwards, D., Halvorson, C., Massie, S., Craig, C., et al.: Overview and characterization of retrievals of temperature, pressure, and atmospheric constituents from the High Resolution Dynamics Limb Sounder (HIRDLS) measurements, Journal of Geophysical Research: Atmospheres, 114, 2009.

Knepp, T. N., Kovilakam, M., Thomason, L., and Miller, S. J.: Characterization of stratospheric particle size distribution uncertainties using SAGE II and SAGE III/ISS extinction spectra, Atmospheric Measurement Techniques, 17, 2025–2054, https://doi.org/10.5194/amt-17-2025-2024, 2024.

Li, Y., Dhomse, S. S., Chipperfield, M. P., Feng, W., Chrysanthou, A., Xia, Y., and Guo, D.: Effects of reanalysis forcing fields on ozone trends and age of air from a chemical transport model, Atmospheric Chemistry and Physics, 22, 10 635–10 656, 2022.

Li, Y., Dhomse, S. S., Chipperfield, M. P., Feng, W., Bian, J., Xia, Y., and Guo, D.: Quantifying stratospheric ozone trends over 1984–2020: a comparison of ordinary and regularized multivariate regression models, Atmospheric Chemistry and Physics, 23, 13 029–13 047, 2023.

Lickley, M., Fletcher, S., Rigby, M., and Solomon, S.: Joint inference of CFC lifetimes and banks suggests previously unidentified emissions, Nature Communications, 12, 2920, 2021.

Millán, L. F., Livesey, N. J., Santee, M. L., and von Clarmann, T.: Characterizing sampling and quality screening biases in infrared and microwave limb sounding, Atmospheric Chemistry and Physics, 18, 4187–4199, https://doi.org/10.5194/acp-18-4187-2018, 2018.

Millán, L. F., Manney, G. L., Boenisch, H., Hegglin, M. I., Hoor, P., Kunkel, D., Leblanc, T., Petropavlovskikh, I., Walker, K., Wargan, K., and Zahn, A.: Multi-parameter dynamical diagnostics for upper tropospheric and lower stratospheric studies, Atmospheric Measurement Techniques, 16, 2957–2988, https://doi.org/10.5194/amt-16-2957-2023, 2023.

Montzka, S., Butler, J., Elkins, J., Thompson, T., Clarke, A., and Lock, L.: Present and future trends in the atmospheric burden of ozone-depleting halogens, Nature, 398, 690–694, 1999.

Montzka, S. A., Dutton, G. S., Yu, P., Ray, E., Portmann, R. W., Daniel, J. S., Kuijpers, L., Hall, B. D., Mondeel, D., Siso, C., et al.: An unexpected and persistent increase in global emissions of ozone-depleting CFC-11, Nature, 557, 413–417, 2018.

Newman, P. A., Lait, L. R., Kramarova, N. A., Coy, L., Frith, S. M., Oman, L. D., and Dhomse, S. S.: Record high march 2024 Arctic total column ozone, Geophysical Research Letters, 51, e2024GL110 924, 2024.

Ramanathan, V., Cicerone, R. J., Singh, H. B., and Kiehl, J. T.: Trace gas trends and their potential role in climate change, Journal of Geophysical Research: Atmospheres, 90, 5547–5566, 1985.

Rigby, M., Park, S., Saito, T., Western, L., Redington, A., Fang, X., Henne, S., Manning, A., Prinn, R., Dutton, G., et al.: Increase in CFC-11 emissions from eastern China based on atmospheric observations, Nature, 569, 546–550, 2019.

Schmidt, M., Bernath, P., Boone, C., Lecours, M., and Steffen, J.: Trends in atmospheric composition between 2004–2023 using version 5 ACE-FTS data, Journal of Quantitative Spectroscopy and Radiative Transfer, 325, 109 088, 2024.

Tegtmeier, S., Hegglin, M. I., Anderson, J., Bourassa, A., Brohede, S., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Kasai, Y., Krüger, K., Kyrölä, E., Lingenfelser, G., Lumpe, J., Nardi, B., Neu, J., Pendlebury, D., Remsberg, E., Rozanov, A., Smith, L., Toohey, M., Urban, J., von Clarmann, T., Walker, K. A., and Wang, R. H. J.: SPARC Data Initiative: A comparison of ozone climatologies from international satellite limb sounders, Journal of Geophysical Research: Atmospheres, 118, 12,229–12,247, https://doi.org/10.1002/2013JD019877, 2013.

Tegtmeier, S., Hegglin, M. I., Anderson, J., Funke, B., Gille, J., Jones, A., Smith, L., von Clarmann, T., and Walker, K. A.: The SPARC Data Initiative: comparisons of CFC-11, CFC-12, HF and SF_6 climatologies from international satellite limb sounders, Earth System Science Data, 8, 61–78, https://doi.org/10.5194/essd-8-61-2016, 2016.

Tibshirani, R.: Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society, Series B (Methodological), 58, 267–288, 1996.

Velders, G. J., Andersen, S. O., Daniel, J. S., Fahey, D. W., and McFarland, M.: The importance of the Montreal Protocol in protecting climate, Proceedings of the National Academy of Sciences, 104, 4814–4819, 2007.

Weber, M., Dhomse, S., Wittrock, F., Richter, A., Sinnhuber, B., and Burrows, J.: Dynamical control of NH and SH winter/spring total ozone from GOME observations in 1995-2002, GEOPHYSICAL RESEARCH LETTERS, 30, https://doi.org/10.1029/2002GL016799, 2003.

WMO: Scientific Assessment of Ozone Depletion: 2010, Global Ozone Research and Monitoring Project, Global Ozone Research and Monitoring Project—Report No. 52, World Meteorological Organization, 2011.

WMO: Scientific Assessment of Ozone Depletion: 2014, Tech. rep., World Meteorological Organization, Global Ozone Research and Monitoring Project, Report No. 55, https://www.wmo.int/pages/prog/arep/gaw/ozone_2014/, 2015.

WMO: Scientific Assessment of Ozone Depletion: 2018, Tech. rep., World Meteorological Organization, Global Ozone Research and Monitoring Project, Report No. 58, https://csl.noaa.gov/assessments/ozone/2018/, 2018.

WMO: Scientific Assessment of Ozone Depletion: 2022, Tech. rep., World Meteorological Organization, Global Ozone Research and Monitoring Project, GAW Report No. 278, https://csl.noaa.gov/assessments/ozone/2022/, 2022.

Zou, H. and Hastie, T.: Regularization and Variable Selection via the Elastic Net, Journal of the Royal Statistical Society, Series B (Statistical Methodology), 67, 301–320, https://doi.org/10.1111/j.1467-9868.2005.00503.x, 2005.