CO2MVS RESEARCH ON SUPPLEMENTARY OBSERVATIONS

D2.5 A prototype for a simplified chemistry scheme to describe observed variations in NO_2 on spatial scales of ~25 km, suitable for global-scale models.

Due date of deliverable	December, 2024
Submission date	December 2024
File Name	CORSO-D2-5-V4
Work Package /Task	WP2/Task 2.5
Organisation Responsible of Deliverable	UEDIN
Author name(s)	Chlöe Schooling, Paul Palmer, Auke Visser
Revision number	V4
Status	Issued
Dissemination Level / location	Public

The CORSO project (grant agreement No 101082194) is funded by the European Union.
Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.

1 Executive Summary

This work has been done as part of WP2, which focuses on the use of co-emitted species to better estimate anthropogenic emissions in the future CO2MVS capacity. The anthropogenic signal from satellite observations of NO_2 is generally much clearer than that of CO_2.

This report contributes to the objectives of WP2 with a methodology developed for a chemistry scheme for light-weight modelling of NO_2. Specifically, random forest regression models were developed to determine the NO_x chemistry net rate of change and the NO_2:NO ratio. These models were trained on GEOS-Chem 0.25°x0.3125° resolution data centred on mainland Europe, for the year 2019. The models perform very well when tested on unseen modelled data for 2019 with emission perturbations applied (R^2>0.95 for NO_x chemistry, R^2>0.99 for NO_2:NO ratio). In addition, a reasonable performance was found when testing on the unseen year, 2021 (R^2>0.79 for NO_x chemistry, R^2>0.92 for NO_2:NO ratio). We find these prediction models can reproduce NO_2 columns with negligible deviation from a GEOS-Chem full-chemistry model output. The model reconstruction error on NO_2 is found to be smaller than the TROPOMI NO_2 column precision in 99.98% of reconstructed data points.

This work lays the foundation for a Ensemble Kalman Filter based NO_x:CO_2 inversions that we will later perform using the regression-based GEOS-Chem models presented here. Additionally, the methodologies described in this report will be applied to the Integrated Forecasting System (IFS) model with an incremental 4D-Var algorithm for use in multi-scale global IFS inversion analysis (D2.7).

Table of Contents

2 Introduction

2.1 Background

To enable the European Union (EU) to move towards a low-carbon economy and implement its commitments under the Paris Agreement, a binding target was set to cut emissions in the EU by at least 40% below 1990 levels by 2030. European Commission (EC) President von der Leyen committed to deepen this target to at least 55% reduction by 2030. This was further consolidated with the release of the Commission's European Green Deal on the 11th of December 2019, setting the targets for the European environment, economy, and society to reach zero net emissions of greenhouse gases in 2050, outlining all needed technological and societal transformations that are aiming at combining prosperity and sustainability. To support EU countries in achieving the targets, the EU and European Commission (EC) recognised the need for an objective way to monitor anthropogenic CO_2 emissions and their evolution over time.

Such a monitoring capacity will deliver consistent and reliable information to support informed policy- and decision-making processes, both at national and European level. To maintain independence in this domain, it is seen as critical that the EU establishes an observation-based operational anthropogenic CO_2 emissions Monitoring and Verification Support (MVS) (CO2MVS) capacity as part of its Copernicus Earth Observation programme.

The CORSO research and innovation project will build on and complement the work of previous projects such as CHE (the CO_2 Human Emissions), and $CoCO_2$ (Copernicus CO_2 service) projects, both led by ECMWF. These projects have already started the ramping-up of the CO2MVS prototype systems, so it can be implemented within the Copernicus Atmosphere Monitoring Service (CAMS) with the aim to be operational by 2026. The CORSO project will further support establishing the new CO2MVS addressing specific research & development questions.

The main objectives of CORSO are to deliver further research activities and outcomes with a focus on the use of supplementary observations, e.g., co-emitted species or auxiliary observations to better separate fossil fuel emissions from the other sources of atmospheric CO_2. CORSO will deliver improved estimates of emission factors/ratios and their uncertainties as well as the capabilities at global and local scale to optimally use observations of co-emitted species to better estimate anthropogenic CO_2 emissions. More broadly, CORSO will also provide clear recommendations to CAMS, ICOS, and WMO about the potential added-value of high-temporal resolution $^{14}CO_2$ (radiocarbon) and APO (atmospheric potential oxygen) observations as tracers for anthropogenic emissions in both global and regional scale inversions and develop coupled land-atmosphere data assimilation in the global CO2MVS system constraining carbon cycle variables with satellite observations of soil moisture, LAI (leaf area index), SIF (solar induced fluorescence), and biomass. Finally, CORSO will provide specific recommendations for the topics above for the operational implementation of the CO2MVS within the Copernicus programme.

2.2 Scope of WP2

The work presented in this report is part of WP2 of CORSO, which deals with “Use of co-emitted species (correlations, improved emission ratios, uncertainties) in data assimilation systems". The aim of WP2 is to improve the use of observations of co-emitted species (NO_2, CO) to better estimate anthropogenic CO_2 emissions in the future CO2MVS capacity. This is based on the recognition that anthropogenic CO_2 emission estimates cannot generally be constrained using CO_2 concentration observations alone, and the detectability of the anthropogenic signal of co-emitted species, typically with much shorter atmospheric lifetimes, is often much better than that of CO_2. For the emission estimation development at local scale, this WP focuses on the development of methods to increase the accuracy of annual CO_2

emission estimates of hot spots, industrial and urban areas by integrating satellite observations of co-emitted species (NO_2 and CO) in data assimilation systems. Since CO_2 satellite observations are temporally sparse (even with the future CO2M constellation), temporal sampling biases are a significant source of uncertainty in annual CO_2 emission estimates of hot spots. Co-emitted species such as CO and NO_2 are and will be available at sub-diurnal temporal coverage from current and future LEO and GEO satellites. They can therefore be used to improve the constraint on the temporal variability of CO_2 emissions and hence for reducing the uncertainty in annual estimates. The local and regional studies will focus on three regions: Europe, Africa, and Southeast Asia.

2.2.1 Scope of this deliverables

Since it is computationally too expensive to conduct full-chemistry simulations in the global IFS-based CO2MVS at the scale of emission plumes, a fast and reliable chemistry scheme needs to be developed that can be used in the lightweight local mass-balance approaches (T2.1) and in the global CO2MVS (T2.4). We develop and test computationally efficient chemistry schemes that can be applied in data-driven emission quantification and low-resolution global models. We parameterise the NO_x net chemical rate of change and the NO_2:NO ratio to improve our ability to exploit atmospheric NO_2 while minimising additional computational overhead from atmospheric chemistry. ECMWF will contribute to this task to link it to the development of the global IFS CO2MVS. This NO_x chemistry scheme will be tested within the global IFS- based CO2MVS and compared with another ML-based surrogate model of the CAMS chemical model under development at ECMWF.

2.2.2 Work performed in this deliverable

The following activities have been conducted in order to achieve the deliverable; they are presented in detail in Section 3:

- Random forest regression models for NO_x chemistry rate and NO_2:NO were trained on GEOS-Chem model runs for a European Domain in 2019.
- Regression models were tested and validated on perturbed model runs for the same period, and for the unseen year 2021.
- An additional methodology for predicting the change in NO_x chemistry under emission perturbations using a scaling-based method is presented and tested using the same modelled data. This is in preparation for upcoming ensemble Kalman filter calculations.
- A new simplified GEOS-Chem forward model was developed to model the transport and emissions of atmospheric NO_x using the presented chemistry schemes to predict the chemistry rates (regression-based GEOS-Chem). The output NO_x columns were compared to the GEOS-Chem full-chemistry model run.
- The NO_2:NO regression model was used to convert these NO_x columns to NO_2 columns (with the TROPOMI averaging kernels applied). This reconstruction was again compared to the GEOS-Chem full-chemistry model run.
- The reconstruction error for this light-weight methodology for modelling NO_2 was compared to the deviation between GEOSChem and TROPOMI, as well as the magnitude of the internal TROPOMI precision values.

2.2.3 Deviations and counter measures

There were no deviations from the original work plan.

2.3 Project partners

CORSO Partners / Collaborators /Laboratories	
EIDGENOSSISCHE MATERIALPRUFUNGS- UND FORSCHUNGSANSTALT	EMPA
EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS	ECMWF
KAMINSKI THOMAS HERBERT	iLab
NEDERLANDSE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK	TNO
UNIVERSITY OF EDINBURGH	UEDIN
UNIVERSITE PAUL SABATIER TOULOUSE III	UT3
WAGENINGEN UNIVERSITY	WU

3 Data and methods

Here, we describe the GEOS-Chem atmospheric transport model used to build our random forest regression models, the satellite column data we use to evaluate our parameterised model of atmospheric NO_x chemistry, and details that describe how we develop our random forest regression models.

Figure 1 shows a schematic that provides an overview of the different steps we use to parameterise NO_x chemistry and NO_2:NO columns and relate them to NO_2 so that can be compared with satellite observations. A random forest regression model, or a scaling-based approach can be used to predict the chemistry rates. The modelled NO_x concentrations are then converted to NO_2 using an additional random forest model. This efficient approach significantly reduces GEOS-Chem's computational cost for forward modelling of NO_2 columns. This is particularly useful for data assimilation, allowing anthropogenic NO_x emission perturbations to be compared with satellite NO_2 observations, such as TROPOMI.

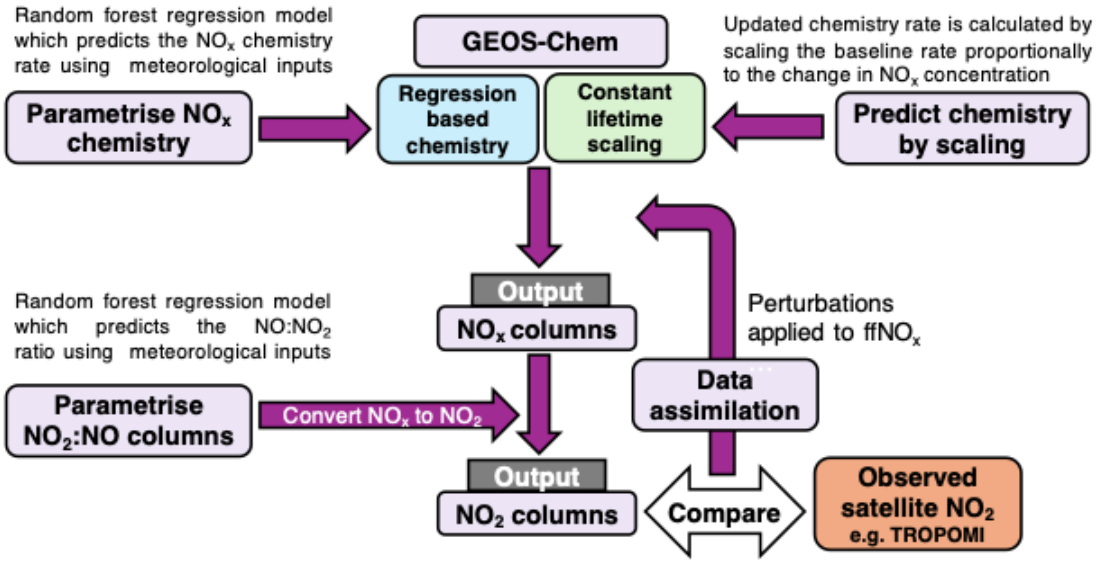


Figure 1: A schematic illustrates how NO_x chemistry parameterization models are integrated into GEOS-Chem for modelling of atmospheric NO_x without a full chemistry scheme.

3.1 GEOS-Chem Atmospheric Chemistry Transport Model

We use version 14.2.2 of the GEOS-Chem atmospheric chemistry transport model to describe the emissions, transport, and chemical production/loss of atmospheric NO_x. For the purpose of our study, we use a nested version of the full chemistry model, centred over mainland Europe (32.75 to 61.25° N, -15 to 40 ° E) with 47 vertical levels, approximately 30 of which fall below the dynamic tropopause. The nested model runs with a horizontal spatial resolution of 0.25°x0.3125°. Initial conditions and lateral boundary conditions to the nested domain were created from a consistent global version of the GEOS-Chem model run at 4°×5°, with three-hourly output fields. We ran the model with a transport timestep of 5 minutes and a chemistry timestep of 10 minutes.

The model is driven by offline meteorology fields from the GEOS Forward Processing (GEOS-FP) product from the Global Modelling and Assimilation Office (GMAO) at NASA Goddard Space Flight Center. GEOS-FP has a native horizontal resolution of 0.25°x0.3125° with 72 vertical pressure levels and 3 hr temporal resolution. To describe the emissions of NO_x, which

differ by year, we used anthropogenic emissions from the Community Emissions Data System (CEDS) version 2 (Hoesly et al., 2018), which provides NO_x emissions in NO_2 equivalent for anthropogenic combustion (industry, energy extraction), and non-combustion sources (agriculture, solvents), including surface transport and shipping. The partitioning of NO_x emissions assumes that 65.2% of NO_x emissions are emitted as NO. Aircraft emissions for NO and NO_2 are taken from the Aviation Emissions Inventory Code (AEIC) (Simone et al., 2013). Pyrogenic emissions of NO are taken from the Global Fire Emissions Database (GFED) version 4.1 (Randerson et al., 2017). In addition, the NO_x emissions from soil and lightning are parameterised within GEOS-Chem (Vinken et al., 2014; Gressent et al., 2016). The NO_x concentration, the NO_x chemical rates of change, and relevant meteorology were output at a temporal resolution of one hour. The chosen meteorological parameters are shown in Table 1. These were selected as they were all found to have a relationship with the net NO_x chemical rate of change.

Parameter	Description	Units
NO_x	Species concentration	molec cm^{-3}
SZA	Solar zenith angle at grid point	degrees
Longitude	Grid point coordinate	degrees-East
Latitude	Grid point coordinate	degrees-North
Altitude	Height above ground level	m
Radiation	Incident short wave radiation	W m^{-2}
Temperature	Atmospheric temperature	K
Humidity	Water vapour mixing ratio	vol vol^{-1}
Wind speed	Wind speed magnitude	m s^{-1}

Table 1. Input parameters used in regression analysis to predict the NO_x chemical net rate of change [molec cm^{-3} s^{-1}] and NO_2:NO ratio.

The model was run for the full year 2019 with baseline (unperturbed) NO_x anthropogenic emissions taken from the CEDs emission inventory. This data was used to train the regression models. To further validate the regression model's performance under varying emissions, additional model runs were conducted with random perturbations applied to anthropogenic NO_x emissions on the order of ±20%. We chose this size of perturbation because a 20% increase in emissions induces changes in NO_2 columns on the same order of magnitude as the difference observed between GEOS-Chem and TROPOMI (as in Fig. 2a). These perturbed runs were performed for 10 days in January, April, July, and October. A model run for the year 2021 was also performed in order to test the regression performance for an unseen time period.

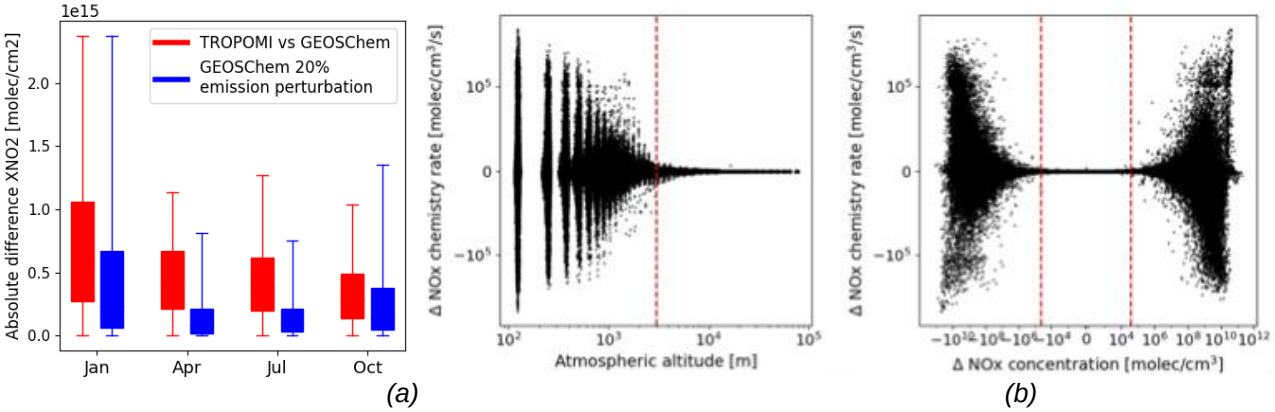


Figure 2: a) Sensitivity testing shows that the impact of 20% emission perturbations on modelled NO2 columns is on the same order as the deviations between GEOS-Chem and TROPOMI. (b) The impact

of emission perturbations on the NO_x chemistry rate becomes small above 3km from the ground, and when the change in NO_x concentration $< 5 \times 10^4$ molec/cm^3.

3.2 Random Forest regression modelling

We trained two random forest regressor models to predict the NO_x net chemical rate of change, and the NO_2:NO ratio. These models were built using the Sci-kit learn python package (Pedregosa et al., 2011). We performed hyperparameter tuning to minimise the computational time of model prediction while maintaining adequate prediction performance (see Fig. A1).

We separately trained both regression models for each month of the year 2019. The models were developed using the NO_x concentration, the spatial location and meteorological variables as input parameters. We then applied a forward selection feature extraction procedure, using mean absolute errors, to further optimise model performance. Based on this procedure, we selected a set of nine features (table 1) for both prediction models. The individual relationship between each of the nine features and the NO_x chemistry rate of change are shown in Fig. A2. We also considered other parameters, including air pressure, air density, the planetary boundary layer height, and the relative mixing ratio of ozone and carbon monoxide (CO), but these were excluded during feature selection.

We trained and tested our NO_x chemistry regression models on model grid points over mainland Europe in the first 3 km above the surface –the region where changes to surface emissions were found to directly influence the atmospheric chemistry, see Fig. 2b. The regression model for the NO_2:NO ratio was predicted for each level in the troposphere and trained on the subset of model data that coincides with the TROPOMI swath (11:30 - 15:30 overpass). The NO_2:NO ratio can be used to convert the concentration:

$$NO_2 = NO_x \frac{NO_2{:}NO}{1+NO_2{:}NO} \quad (1)$$

We test both models on unseen data from model runs that include ±20% emission perturbations similar to those used in an ensemble Kalman filter (Feng et al., 2009, 2023), as well as from an unseen year, 2021. To assess the performance of the regression models, we used the coefficient of determination, R^2, the mean absolute error (MAE), and the mean bias. These are defined by the following equations, where y_i are true values, $\hat{y}_i$ are predicted values, $\bar{y}$ is the mean of the true values, and N is the number of datapoints:

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

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (3)$$

$$Mean\ bias = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i) \quad (4)$$

3.3 Constant lifetime scaling

In an alternative formulation, we apply the assumption that the lifetime of atmospheric NO_x remains constant under stable meteorological conditions. Hence, if a full chemistry model run is available for a baseline emission scenario, the chemistry rates for perturbed scenarios can be calculated by scaling the original rate according to the proportional change in NO_x concentration.

This approach serves as an alternative to using regression models for predicting the chemistry rates. The atmospheric lifetime, τ of NO_x is given by:

$$\tau = \frac{NO_x}{R_{NO_x}} \quad (3)$$

where NO_x denotes the combined NO and NO_2 species concentrations [molec cm^{-3}] and R_{NOx} is the chemical rate of change [molec cm^{-3} s^{-1}] that describes the net loss, which accounts for the balance between its chemical production (e.g., from reactions involving NO or NO_2 precursors) and its chemical loss processes (e.g., reactions forming reservoirs like HNO_3 or NO_y species). Note that when NO_x experiences net chemical production, the atmospheric lifetime becomes negative. The benefit of looking at chemical lifetime, rather than the net rate of change, is that the quantity is largely independent of species concentration. This independence allows for a more stable understanding of the NO_x chemistry, irrespective of fluctuations in its concentration caused by emission changes.

We found that while the influence of ±20% emission perturbations cause clear changes to the NO_x chemical net rate of change, the resulting changes to atmospheric lifetime are considerably smaller (see Fig. A3). This result suggests that the chemical lifetime is driven by the meteorology and location in the model but is less sensitive to changing concentrations of NO_x. The unperturbed model run provides NO_x concentrations and rates of change at a 1-hour temporal resolution, allowing the chemical rate of change to be updated every hour under the assumption of an unchanged chemical lifetime. The new rate of change can be determined using the NO_x lifetime, τ, and the local NO_x concentration:

$$R_{NO_x}(x, y, z, t) = \frac{NO_x(x,y,z,t)}{\tau(x,y,z,t)} \quad (4)$$

For this method, an initial unperturbed full-chemistry model run must be employed to determine the NO_x chemical lifetime $\tau(x, y, z, t)$ for each grid-point and time-point for the spatial and temporal region of interest. Then for any further perturbed model runs, the chemistry rates can be determined without the need of an integrated chemistry scheme, thereby saving considerable computational time. The updated chemistry rates are then simply scaled by the ratio of the new NO_x concentration to the original NO_x concentration; so, if the concentration doubles, then we assume a doubling in the net chemical rate of change.

3.4 Regression-based chemistry atmospheric transport modelling

For this study, we added the NO_x species to the GEOS-Chem tagged carbon model, CO_2, CO, methane, and carbonyl sulphide, in which individual tagged tracers track contributions of these trace gases from geographical regions and/or natural and human- driven fluxes. This model does not include an integrated chemistry scheme and therefore the NO_x species chemical rate of change is determined using the NO_x chemistry regression model. Going forward, we refer to this model as the regression-based atmospheric chemistry model (shown in Fig. 1).

We performed a full-chemistry model run with emission perturbations to evaluate the impact of emission changes on NO_x chemistry, and later to assess the performance of our regression model in predicting the effects of emission changes. An analysis of how the emission-driven changes in chemistry rate varied with the atmospheric altitude as well as the change in NO_x concentration is shown in Fig. 2b. The net rate of change in NO_x chemistry showed minimal variability at altitudes below 3 km, where the chemistry change was less than 9×10^3 molec/cm^3/s. Additionally, minimal variability in atmospheric chemistry was observed when the absolute change in NO_x concentration was less than 5×10^4 molec/cm^3, which corresponds

to a chemistry change of less than 2×10^3 molec/cm^3/s. Based on these findings, we set a condition to update the NO_x net chemical rate of change using the unperturbed full-chemistry outputs for altitudes above 3 km and for regions where the change in NO_x concentration is less than 5×10^4 molec/cm^3. For all other regions, the chemistry regression model is used to predict the new rate of change.

We also used the constant lifetime scaling method (see above) to predict the new rate of change. Looking to Fig. 1 we can see that this methodology provides an alternative approach to the regression-based atmospheric chemistry model for modelling NO_x columns. Throughout this report we will compare the results of the regression-based chemistry scheme and the constant lifetime scaling based approach.

We ran the model for 10 days in January, April, July, and October which provided contrasting seasonal conditions to test the model. For each run, we use the ±20% perturbed anthropogenic NO_x emission sets. To evaluate the veracity of the NO_x column model outputs for the regression-based chemistry model and for the constant lifetime scaling model, we compare them with the full-chemistry model outputs. We use our NO_2:NO ratio regression model to convert NO_x results from our atmospheric chemistry regression model to NO_2 columns, sampled at the time and location of TROPOMI data, so they can be compared with TROPOMI NO_2 column data.

3.5 TROPOMI Satellite Column Observations of NO_2

We use TROPOMI NO_2 tropospheric columns to compare with the GEOSChem model output. TROPOMI was launched in 2017 in a Sun-synchronous orbit with a local equatorial overpass time of 13:30. It has a swath width of 2600 km and a ground pixel of 7×7 km^2 in the nadir. Due to the width of the swath, the 13:30 overpass time corresponds to data captured with local solar time (LST) ranging from 11:30 and 15:30 in the highest latitude regions of the European domain. We only used data with a quality flag ≥ 0.75, filtering out data affected by elevated cloud cover, aerosol loading, and larger solar and viewing zenith angles. We analysed TROPOMI data for 10 days in January, April, July, and October 2019.

For our study, we regridded TROPOMI data to our $0.25^\circ\times0.3125^\circ$ GEOS-Chem model grid. To enable a comparison between TROPOMI and GEOS-Chem, we sampled the model at the location and time of each TROPOMI observation. We applied scene-dependent TROPOMI averaging kernels, describing the instrument sensitivity to changes in atmospheric NO_2, to the corresponding model NO_2 profiles.

4 Results and discussion

Here, we report the model performance of our atmospheric chemistry prediction models for NO_x and the accompanying regression model for the NO_2:NO ratio that enables us to convert NO_x columns to NO_2 columns observed by satellites. We assess the fidelity of our results from these models using the full-chemistry version of GEOS-Chem and evaluate our results using TROPOMI NO_2 column data.

4.1 Performance of atmospheric chemistry regression models for NO_x

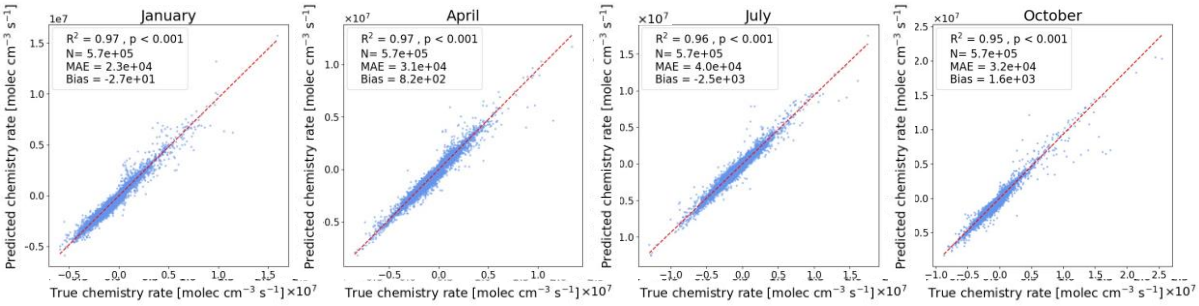
4.1.1 NO_x chemistry random forest

Fig. 3a shows that the NO_x chemistry random forest model has a very good performance at reproducing results from the full-chemistry version of GEOS-Chem with 20% emission perturbations applied to the testing dataset for the four months we study in 2019. The model performance R^2 values are 0.97, 0.97, 0.96 and, 0.95 for January, April, July, and October 2019, respectively. The MAE values are largest in July (4×10^4 molec/cm^3/s) and smallest in January (2.3×10^4 molec/cm^3/s), reflecting the range in magnitude of chemistry rates that increase during summer months over Europe.

We also tested our regression-based atmospheric chemistry model with model data from 2021 (Fig. A4). As expected, the regression model performance has less skill in reproducing data with new meterological conditions that have not been used for training. In this case, the MAE values are higher by a factor of 1.3-1.8 compared with the overall performance comparison shown in Fig. 4. Nevertheless, the model still shows substantial skill despite substantial differences in anthropogenic emissions between 2019 and 2021 due to COVID-19. Specifically, NO_x emissions were found to decrease by 18-24% during lockdown periods (Miyazaki et al., 2021) leading to a mean observed reduction in NO_2 of 29% (Cooper et al., 2022).

4.1.2 NO_x chemistry prediction using constant lifetime scaling NO_x chemistry random forest

Fig. 3b shows results from using our alternative atmospheric chemistry regression NO_x model that employs a constant atmospheric lifetime scaling approach (eq. 4). The resulting model performance is a significant improvement above the other regression model for all four study months. Using our scaling approach, we found consistent values of R^2 = 1.0 and MAE values that are approximately 2-3 times smaller than the other regression model. As with the other regression model, the size of the error is scaled by the seasonal changes in chemistry rates.



(a)

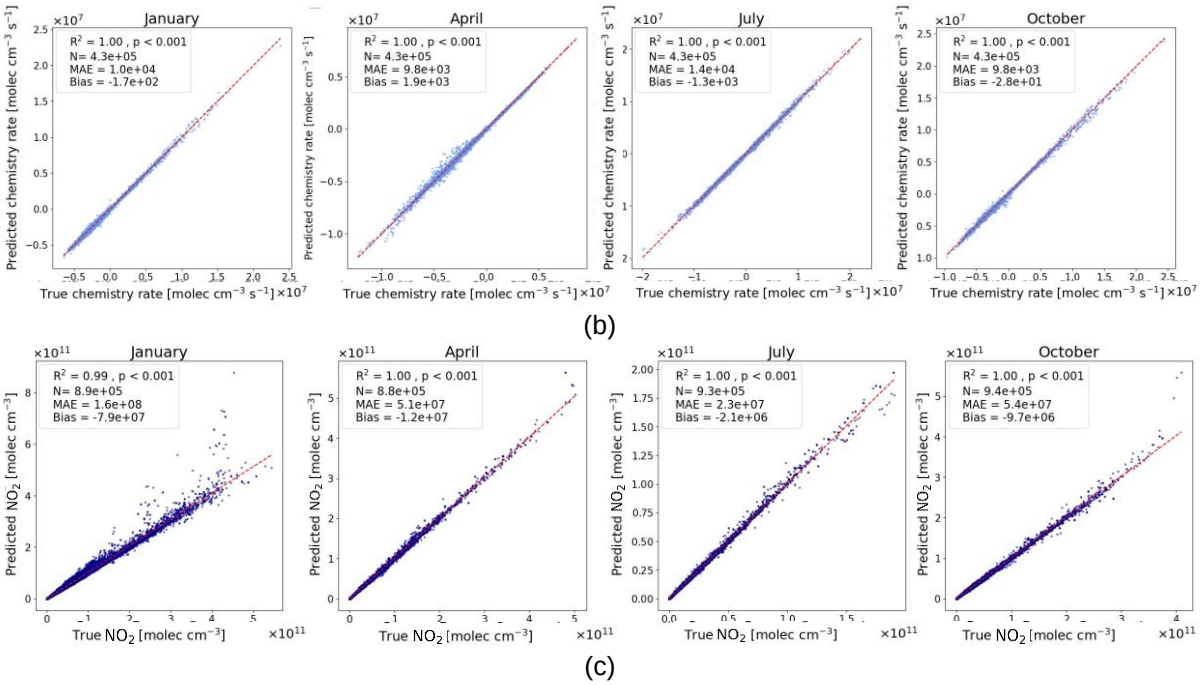


Figure 3: Actual versus predicted scatter plots for (a) the random forest regression model for predicting the NO_x chemistry rate, (b) the constant lifetime scaling for reconstructing the NO_x chemistry rate using an unperturbed chemistry dataset, (c) the reconstruction of NO_2 from NO_x using the random forest regression model for predicting the NO_2:NO ratio.

While this approach shows extremely encouraging abilities to determine NO_x chemistry rates, its effectiveness relies on having a full-chemistry model run available for at least one set of emission inputs. Consequently, this approach is particularly useful for emission perturbation studies, for which numerous emission distribution scenarios might be needed for model inversion work. In this case, the full-chemistry model would only need to be run once for the given time period of interest. However, we cannot predict the NO_x chemistry using this method for a previously unmodelled time period.

4.1.3 NO:NO_2 ratio regression model

We find the random forest regression model to predict NO_2:NO ratios also demonstrates significant performance. The predicted ratio is used to convert NO_x concentrations to NO_2 concentrations (eq. 1). Fig. 3c shows that the regression model can reproduce "true" NO_2 values from the full-chemistry of the GEOS-Chem model, with values of R^2 of 1.0; the exception is January when R^2 = 0.99.

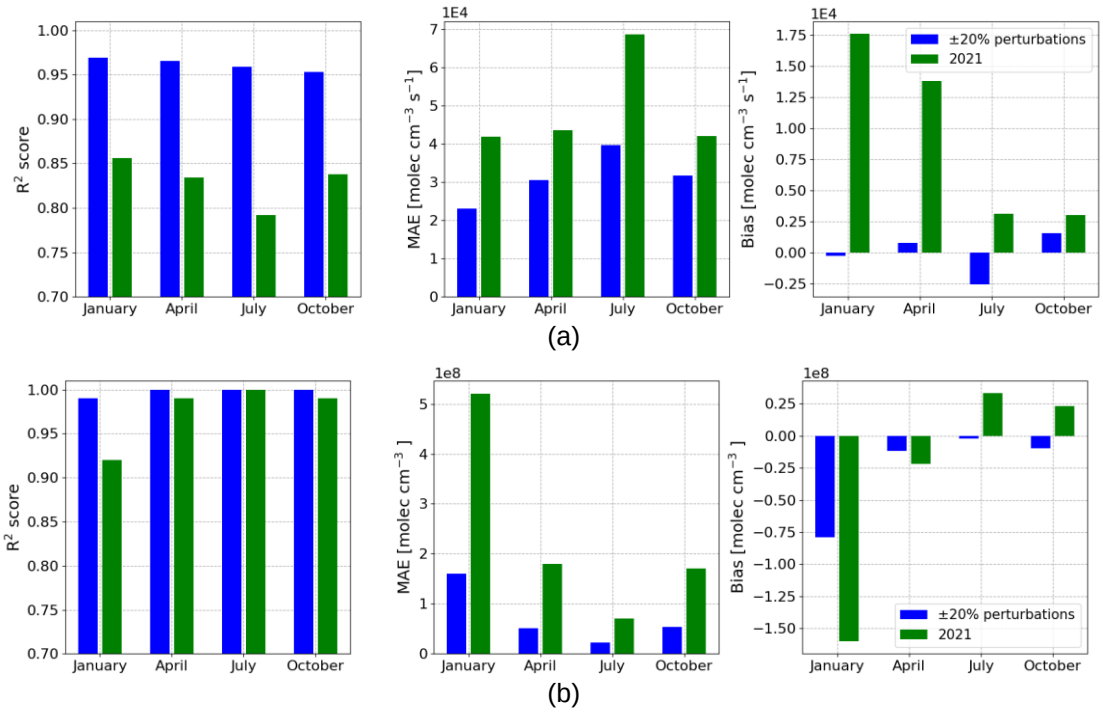


Figure 4: Regression model prediction performance compared when tested on a 20% perturbed model run for 2019 and an unseen year, 2021. (a) Shows the NO_x chemistry regression model performance comparisons and (b) shows the NO_2 prediction performance using the NO_2:NO regression model.

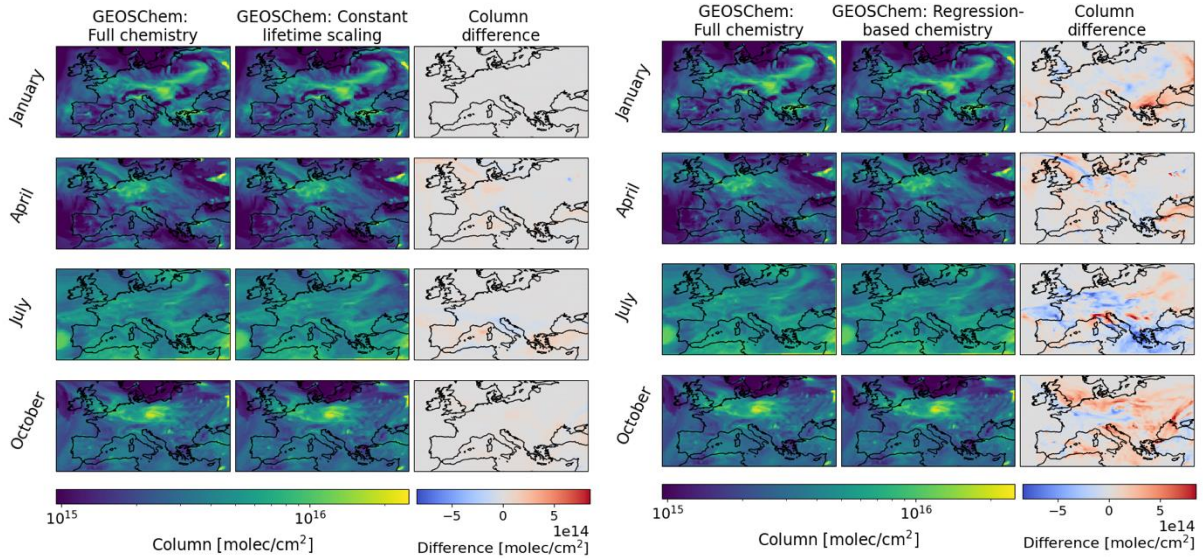
Generally, the model performance is better during summer months and worse in winter months, with MAE values an order of magnitude smaller in July compared to January. This is partly due to NO_2 concentrations increasing during colder months due to increased combustion and longer nights, and because we find that NO_2:NO ratios become increasingly hard to determine at higher solar zenith angles, typically experienced over Europe during daytime through winter months. We also examine the performance of this regression model using data from the unseen year 2021. As with the atmospheric chemistry regression model, described above, the performance was good but worse than for 2019 in which data was used to train the model. The MAE reduced by a factor of 3.25, 3.52, 3.04, and 3.14 for January, April, July, and October respectively. We found the R^2 performance reduced most for January from 0.99 to 0.92, During April and October R^2 reduced from 1.0 to 0.99, while R^2 1.0 was maintained in July.

4.2 Atmospheric modelling of NO_x

Fig. 5 shows the NO_x column reconstruction for the two regression models used to describe the NO_x chemistry rates from the full-chemistry version of the GEOS-Chem model. From a visual inspection, there are no obvious differences in the spatial distribution of the NO_x columns reconstructed using both the regression-based chemistry model and the constant lifetime scaling model. However, when mapping the differences, there are areas of deviation from the full-chemistry model. Broadly, this deviation is significantly smaller when we use the scaling-based model compared to the regression-based. In addition, the error accumulation in January is notably smaller than in other months.

Fig. 6 shows the temporal variation in the reconstruction error. The range, IQR, and median values are shown in 6a and the mean absolute percentage error (MAPE) is shown in 6b. For the regression-based chemistry method the range in deviation peaks at up to 3×10^{14} molec/cm^2 in January, 5×10^{14} molec/cm^2 in April and 6×10^{14} molec/cm^2 in July and October.

This is reflected in maximum MAPE values of 2.8%, 9.7%, 8.9%, and 9.3% for the four months, respectively. On the whole, the MAPE reduces through time, with final deviation values of 1.7%, 3.4%, 2.0%, and 4.8% after the full 10-day run.



(a) (b)

Figure 5: The modelled NO_x columns after a 10-day model run with ±20% emission perturbations. NO_x columns are compared for the GEOS-Chem full-chemistry model and (a) NO_x columns are simulated using the regression-based chemistry method and (b) using the constant lifetime scaling method.

Reconstruction errors for the constant lifetime scaling model show much smaller errors, particularly, with MAPE < 0.2% throughout the 10-day run. This is driven by the smaller impact that emission perturbations have on the NO_x chemistry in January as shown by Fig. A3. In particular, the lifetime of NO_x is relatively unchanged between the unperturbed and perturbed model runs. This reduced impact in January is likely due to the slower rate of photochemical reactions in the winter months and increased atmospheric stability at lower temperatures. The other months do see a more prominent deviation of up to a maximum of 4×1014 molec/cm^2, with peak MAPE values of 6.6%, 5.7%, and 4.5%, for April, July, and October, respectively. As with the regression-based model outputs, here the MAPE also generally decreases through time with final deviation values of 0.1%, 1.1%, 0.2%, and 0.3% for each month, respectively. Interestingly, while the range and IQR are relatively stable throughout the run when using the regression-based reconstruction, these quantities decrease considerably with time when we use the scaling-based reconstruction.

The reconstruction error has a small diurnal cycle, peaking in the morning and to a lesser extent in the evening, reflecting the diurnal cycle of NO_x chemistry. Overall, the absolute model error for both the regression-based and scaling-based methods peaks after the first day and then gradually reduce, plateauing by ≃day 6. It is encouraging that there is no accumulation of error through time, suggesting this approach would be suitable for studies longer than for ten days. It is clear that the optimal reconstruction performance is found when using the scaling-based method, but as we already note there are limitations to this method. The regression-based approach still provides excellent reconstruction performance for our purposes.

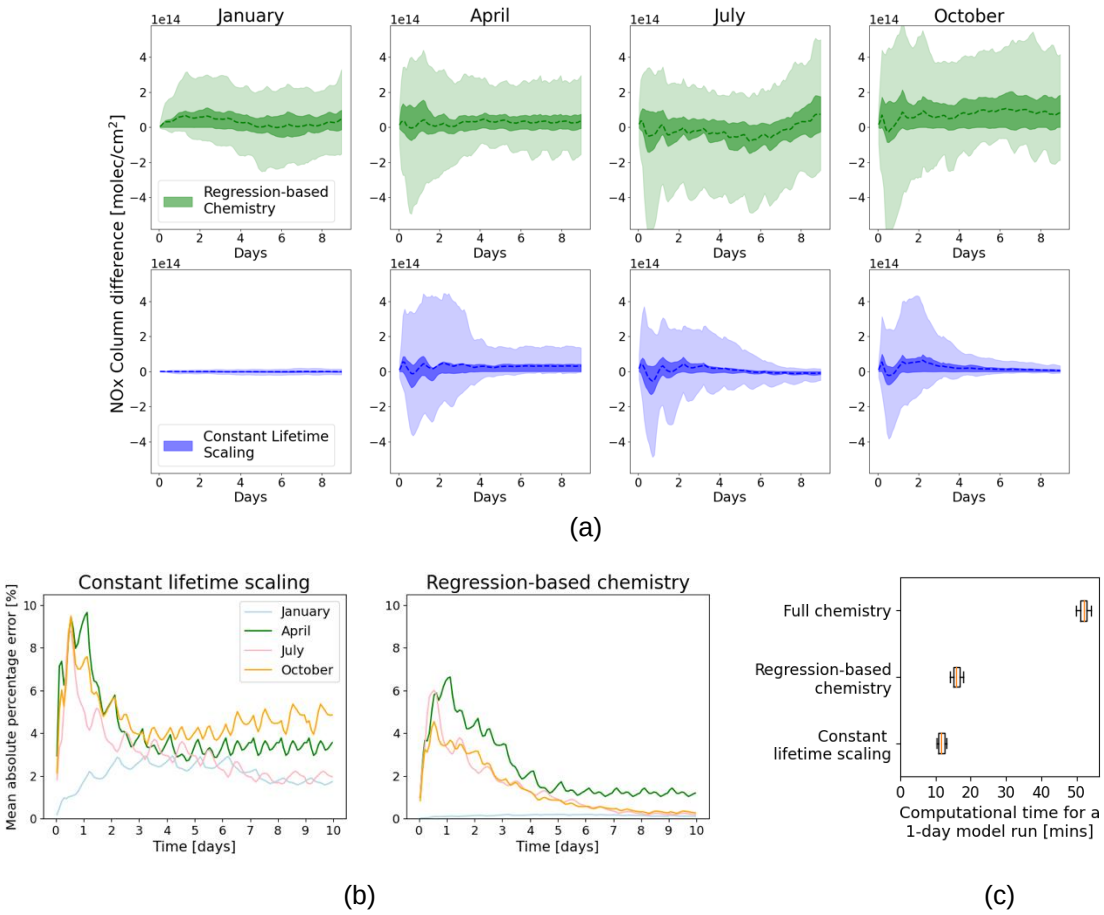


Figure 6: Comparison of the temporal variation in NO_x column reconstruction for the regression-based and scaling-based model. (a) The median (dashed line), IQR (light-shaded region) and range (dark-shaded region) of the NO_x column reconstruction error over the 10-day runs. (b) The mean absolute percentage error over the 10-day runs. (d) Shows the reduction in computational time when modelling atmospheric NO_x using each of our chemistry prediction methods compared to running with the full-chemistry model.

Substantial computational time is saved when we employ these regression methods to model atmospheric NO_x. Fig. 6c shows the time taken for each model to perform a 1-day model run. This was calculated as the mean average for the model to run for a single day out of the 10 days run for each of the four months, repeated for 3 model runs. Clearly, the full-chemistry model takes the longest, with a mean of 52 minutes per day for our nested model over Europe. The regression-based chemistry model is significantly faster with a mean of 16 minutes (3.25 times improvement), while the constant lifetime scaling method is even fast, with a mean of 12 minutes (4.3 times improvement). It is important to note that there will be some variation in these model times depending on the relative loading experienced by the computer system used. In order to perform regression-based chemistry modelling, each model must be trained, which requires approximately 18 hours per monthly model. In contrast, applying the constant lifetime scaling method necessitates an initial full-chemistry model run, taking about 26 hours for a single month. While training regression models also relies on full-chemistry model runs, the advantage is that the trained models can be applied to unseen time periods, saving time. This flexibility is not available when calculating chemistry rates using the scaling method.

4.3 NO_2 column reconstruction

Finally, we assess the capability of our NO_2:NO regression model, convolved with TROPOMI instrument averaging kernels, to reproduce observation column distributions of NO_2 from TROPOMI. The absolute differences in NO_2 columns between GEOS-Chem full-chemistry and the GEOS-Chem regression-based and scaling-based models are compared to the absolute difference in TROPOMI NO_2 and GEOS-Chem full-chemistry, as well as to the magnitude of the TROPOMI NO_2 column precision data. This is presented in Fig. 7, compared for 8 days in January, April, July, and October.

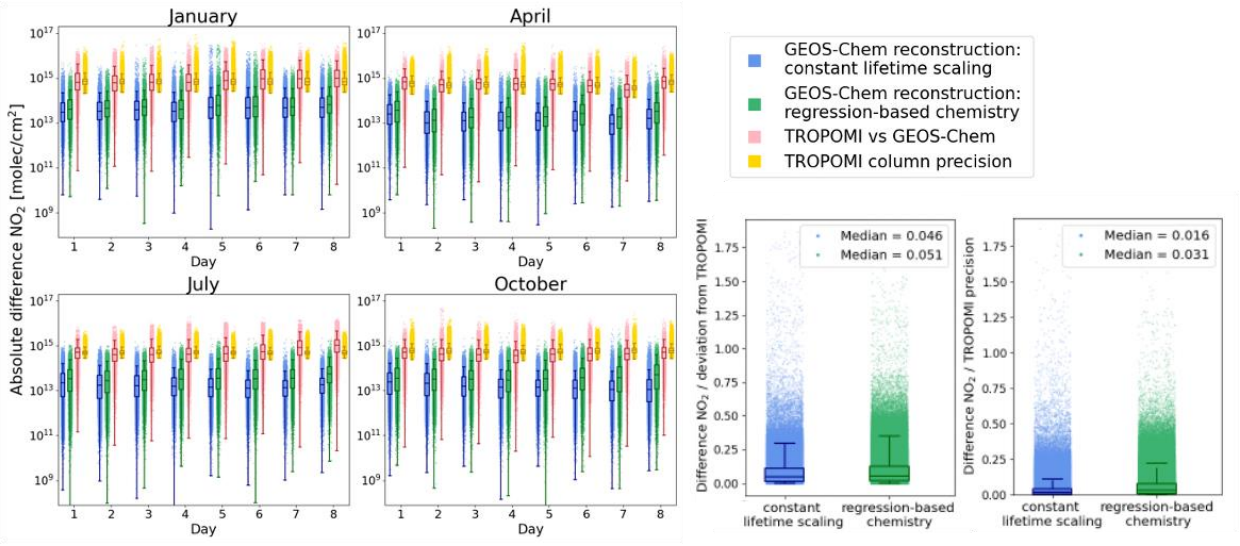


Figure 7: The absolute difference in NO_2 from GEOS-Chem full-chemistry with the constant lifetime scaling based model (blue), with the regression-based chemistry model (green), with TROPOMI NO_2 (red), as well as TROPOMI NO_2 tropospheric column precision values (yellow). In addition, the normalised NO_2 differences are calculated by normalising the reconstructed model deviation by the absolute deviation between GEOSChem and TROPOMI, as well as by the TROPOMI column precision.

We find comparable NO_2 reconstruction errors for the four months we study. Earlier, with the NO_x reconstruction, we found that the error was smaller for January than the other months (Fig. 3a and 3b), however, the higher error from the January NO_2:NO regression model (Fig. 3c offsets this advantage, ultimately bringing the overall reconstruction error for all months to a comparable level. We find something similar when we compare our NO_2 reconstructions based on the scaling- based and regression-based methods, with a comparable magnitude of absolute error between the two. When we compare the difference between GEOS-Chem and TROPOMI NO_2 columns, we find that the regression model NO_2 reconstruction errors are much smaller than the estimated precision values for the data. This provides confidence that our model reconstruction performance is robust enough for use in inversion work as well as for general comparisons with TROPOMI or other observational data. See Appendix B for a more detailed analysis on the difference between modelled column NO_2 and observed TROPOMI data.

Fig. 7 shows that the median NO_2 column model reconstruction errors are about 5% of the actual deviation from TROPOMI – 4.6% and 5.1% for the scaling and regression-based approaches, respectively. Similarly, these construction errors for the scaling and regression-based approaches represent a median value of 1.6%, and 3.1% of the TROPOMI precision values. Looking at all the reconstructed data points for both methods for all the days modelled in our four months of interest, we found that 99.98% of the data gave a reconstruction error smaller than the corresponding TROPOMI column precision.

4.4 Prospects for implementation in IFS

Looking ahead to the implementation of parametrised NO_x chemistry for CO_2 emission inversions with the IFS model, we compared net NO_x rates of change between GEOS-Chem and the IFS. This analysis spans two months (January and July) of 2021 that represent two distinct photo-chemical regimes in Europe. We find that IFS and GEOS-Chem generally agree on the median diurnal variability in the surface model layer, shown in Fig. 8. Nocturnal NO_x removal in January is higher than during daytime in both models, although GEOS-Chem simulates larger nighttime NO_x removal than the IFS. In July, NO_x removal peaks in the early morning before dropping to lower afternoon values in both models, but IFS consistently simulates larger chemical NO_x removal by ~0.25×10^6 molec/cm^3/s. The differences in absolute NO_x rates of change between GEOS-Chem and IFS may be caused by differences in chemistry schemes, and different surface layer heights in the models (20 m in IFS and 60 m in GEOS-Chem). Based on this analysis, we conclude that NO_x chemistry in IFS can be parametrised using similar driving variables as for GEOS-Chem, but that they need to be adjusted to capture absolute values of the IFS-simulated NO_x rate of change.

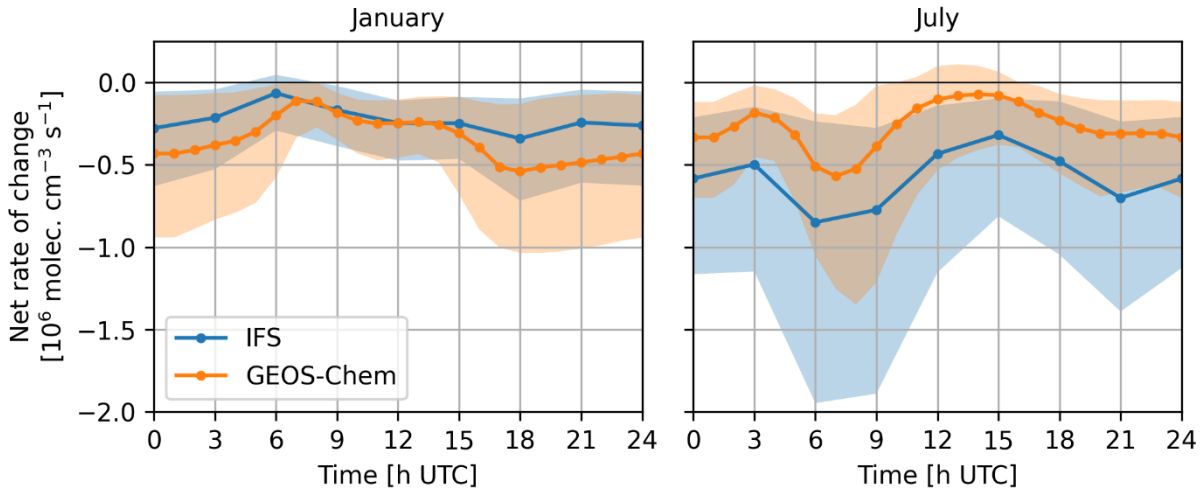


Figure 8: Surface-layer net NO_x rate of change in IFS (blue) and GEOS-Chem (orange) shown for January (left panel) and July (right panel) 2021 in Europe (analysed domain: 15 °W – 40 °E, 32.75 °N – 61.25 °N). Lines and shaded areas display the median and inter-quartile range, respectively.

5 Conclusion

We have demonstrated that the NO_x chemistry rates and NO_2:NO ratio described by a leading 3-D atmospheric chemistry model can be reproduced using random forest-based regression models using NO_x concentrations, the spatial location, and meteorological variables as input parameters. The models perform successfully through all months of 2019 with $R^2 > 0.95$ for predicting NOx chemistry rates and R2 > 0.99 for predicting the corresponding NO2:NO concentration ratios. We also show that these models maintain some prediction capability when tested on model outputs from an unseen year (2021) with contrasting environment conditions. Overall, the prediction power when testing on unseen meteorology increases the MAE by an average factor of 1.5 for the chemistry rate prediction, and by an average factor of 3.3 for the NOx to NO2 conversion performance. Additional work will focus on assessing the impact on modelled NO2 column reconstruction when using regression models trained on meteorological periods excluded from the training data.

We have also demonstrated that the atmospheric lifetime of NO_x is stable against varying emissions, particularly in winter months. From this, we have demonstrated that it is also possible to predict updated NO_x chemistry rates of change as a result of emission perturbations, with knowledge of NO_x chemistry from an initial unperturbed model run. This scaling-based approach has impressive prediction performance with R^2 =1.0.

We have developed two viable methodologies to model atmospheric NO_x in a more computationally efficient way than using the GEOS-Chem 3-D model. The regression-based chemistry method has the advantage of not requiring prior knowledge of the NO_x lifetimes for a baseline model run and reduces the computational time by a factor of 3.25. The lifetime scaling-based approach reduces the model run time slightly further by a factor of 4.3, but a baseline full-chemistry model run is required. This scaling-based approach has smaller model reconstruction errors, but generally both approaches have reconstruction errors smaller than the TROPOMI precision values for 99.98% of the reconstructed data (399,502 points). Both methodologies will be applied in deliverable D2.7, where we will apply model inversions using IFS with an incremental 4D-Var algorithm; and using GEOS-Chem with an ensemble Kalman filter (EnKF) approach. In particular, the 4D-var algorithm involves running an initial full-chemistry simulation, therefore the constant lifetime scaling approach could be efficiently applied in this case.

The regression model development work will be expanded by testing, and if necessary, applying additional training using IFS model outputs. Furthermore, the models will be extended to a global scale and tested on additional years and regions of interest.

Our study provides confidence in random forest models being used to describe NO_x chemistry to a sufficient accuracy for them to play an important role in NO_x:CO_2 inversion methods to improve $ffCO_2$ estimates. This unlocks a more efficient method to infer NO_x from which to infer $ffCO_2$ via emission estimates (Berezin et al., 2013; Lopez et al., 2013; Goldberg et al., 2019; Super et al., 2020). Results from our study are particularly timely with the launch in the next few years of the Copernicus Anthropogenic Carbon Dioxide Monitoring constellation (CO2M) that include column measurements of CO_2 and NO_2. Our future work will implement a combined a joint NO_x:CO_2 model inversion to constrain geographically resolved $ffCO_2$ estimates that will support the development of measurement, reporting and verification systems, contributing to the fundamental aim of CORSO.

6 References

Balsamo, G., Engelen, R., Thiemert, D., Agusti-Panareda, A., Bousserez, N., Broquet, G., Brunner, D., Buchwitz, M., Chevallier, F., Choulga, M., et al.: The CO2 Human Emissions (CHE) Project: First steps towards a European operational capacity to monitor anthropogenic CO2 emissions, *Frontiers in Remote Sensing*, **2**, 707 247, 2021.

Berezin, E. V., Konovalov, I. B., Ciais, P., Richter, A., Tao, S., Janssens-Maenhout, G., Beekmann, M., and Schulze, E.-D.: Multiannual changes of CO2 emissions in China: indirect estimates derived from satellite measurements of tropospheric NO2 columns, *Atmospheric Chemistry and Physics*, **13**, 9415–9438, https://doi.org/10.5194/acp-13-9415-2013, 2013.

Cooper, M. J., Martin, R. V., Hammer, M. S., Levelt, P. F., Veefkind, P., Lamsal, L. N., Krotkov, N. A., Brook, J. R., and McLinden, C. A.: Global fine-scale changes in ambient NO2 during COVID-19 lockdowns, *Nature*, **601**, 380–387, https://doi.org/10.1038/s41586-021-04229-0, 2022.

Feng, L., Palmer, P. I., Bösch, H., and Dance, S.: Estimating surface CO2 fluxes from space-borne CO2 dry air mole fraction observations using an ensemble Kalman Filter, *Atmospheric Chemistry and Physics*, **9**, 2619–2633, https://doi.org/10.5194/acp-9-2619-2009, 2009.

Feng, L., Palmer, P. I., Parker, R. J., Lunt, M. F., and Bösch, H.: Methane emissions are predominantly responsible for record-breaking atmospheric methane growth rates in 2020 and 2021, *Atmospheric Chemistry and Physics*, **23**, 4863–4880, https://doi.org/10.5194/acp-23-4863-2023, 2023.

Goldberg, D. L., Lu, Z., Oda, T., Lamsal, L. N., Liu, F., Griffin, D., McLinden, C. A., Krotkov, N. A., Duncan, B. N., and Streets, D. G.: Exploiting OMI NO2 satellite observations to infer fossil-fuel CO2 emissions from U.S. megacities, *The Science of the Total Environment*, **695**, 133805, 2019.

Gressent, A., Sauvage, B., Cariolle, D., Evans, M., Leriche, M., Mari, C., and Thouret, V.: Modeling lightning-NO_x chemistry on a sub-grid scale in a global chemical transport model, *Atmospheric Chemistry and Physics*, **16**, 5867–5889, https://doi.org/10.5194/acp-16-5867-2016, 2016.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), *Geoscientific Model Development*, **11**, 369–408, https://doi.org/10.5194/gmd-11-369-2018, 2018.

Jacob, D.: *Introduction to atmospheric chemistry*, Princeton University Press, Princeton, NJ, 1999.

Kemball-Cook, S., Yarwood, G., Johnson, J., Dornblaser, B., and Estes, M.: Evaluating NO_x emission inventories for regulatory air quality modeling using satellite and air quality model data, *Atmospheric Environment*, 117, 1–8, 2015.

Liu, X., Ou, J., Wang, S., Li, X., Yan, Y., Jiao, L., and Liu, Y.: Estimating spatiotemporal variations of city-level energy-related CO2 emissions: An improved disaggregating model based on vegetation-adjusted nighttime light data, *Journal of Cleaner Production*, 177, 101–114, 2018.

Lopez, M., Schmidt, M., Delmotte, M., Colomb, A., Gros, V., Janssen, C., Lehman, S. J., Mondelain, D., Perrussel, O., Ramonet, M., Xueref-Remy, I., and Bousquet, P.: CO, NO_x and 1313CO22 as tracers for fossil fuel CO22: results from a pilot study in Paris during winter 2010, *Atmospheric Chemistry and Physics*, 13, 7343–7358, 2013.

Meijer, H., Smid, H., Perez, E., and Keizer, M.: Isotopic characterisation of anthropogenic CO2 emissions using isotopic and radiocarbon analysis, *Physics and Chemistry of the Earth*, 21, 483–487, 1996.

Miyazaki, K., Bowman, K., Sekiya, T., Takigawa, M., Neu, J. L., Sudo, K., Osterman, G., and Eskes, H.: Global tropospheric ozone responses to reduced NO_x emissions linked to the COVID-19 worldwide lockdowns, *Science Advances*, 7, https://doi.org/10.1126/sciadv.abf7460, 2021.

Napelenok, S. L., Pinder, R. W., Gilliland, A. B., Marin, R. V., Miranda, A. I., Borrego, C., and Borrego, C.: Developing a method for resolving NO_x emission inventory biases using discrete Kalman filter inversion, direct sensitivities, and satellite-based NO2 columns, in: *Air Pollution Modeling and Its Application XIX*, NATO Science for Peace and Security Series Series C: Environmental Security, pp. 322–330, Springer Netherlands, Dordrecht, ISBN 9781402084522, 2008.

Nayagam, L., Maksyutov, S., Oda, T., Janardanan, R., Trisolino, P., Zeng, J., Kaiser, J. W., and Matsunaga, T.: A top-down estimation of subnational CO2 budget using a global high-resolution inverse model with data from regional surface networks, *Environmental Research Letters*, 19, 014031, https://doi.org/10.1088/1748-9326/ad0f74, 2023.

Nguyen, D.-H., Lin, C., Vu, C.-T., Cheruiyot, N. K., Nguyen, M. K., Le, T. H., Lukkhasorn, W., Vo, T.-D.-H., and Bui, X.-T.: Tropospheric ozone and NO_x: A review of worldwide variation and meteorological influences, *Environmental Technology Innovation*, 28, 102809, https://doi.org/10.1016/j.eti.2022.102809, 2022.

Oda, T., Feng, L., Palmer, P. I., Baker, D. F., and Ott, L. E.: Assumptions about prior fossil fuel inventories impact our ability to estimate posterior net CO2 fluxes that are needed for verifying national inventories, *Environmental Research Letters*, 18, 124030, https://doi.org/10.1088/1748-9326/ad059b, 2023.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, 12, 2825–2830, 2011.

Randerson, J., Van Der Werf, G., Giglio, L., Collatz, G., and Kasibhatla, P.: Global Fire Emissions Database, Version 4.1 (GFEDv4), https://doi.org/10.3334/ORNLDAAC/1293, 2017.

Shu, Y. and Lam, N. S.: Spatial disaggregation of carbon dioxide emissions from road traffic based on multiple linear regression model, *Atmospheric Environment*, 45, 634–640, 2011.

Simone, N., Stettler, M., Eastham, S., and Barrett, S.: Aviation Emissions Inventory Code (AEIC), https://doi.org/10.5281/ZENODO.6461767, 2013.

Super, I., van der Gon, H. A. C. D., van der Molen, M. K., Dellaert, S. N. C., and Peters, W.: Optimizing a dynamic fossil fuel CO2 emission model with CTDAS (CarbonTracker Data Assimilation Shell, v1.0) for an urban area using atmospheric observations of CO2, CO, NO_x, and SO2, *Geoscientific Model Development*, 13, 2695–2721, 2020.

Super, I., Scarpelli, T., Droste, A., and Palmer, P. I.: Improved definition of prior uncertainties in CO2 and CO fossil fuel fluxes and its impact on multi-species inversion with GEOS-Chem (v12.5), *Geoscientific Model Development*, 17, 7263–7284, https://doi.org/10.5194/gmd-17-7263-2024, 2024.

Vinken, G. C. M., Boersma, K. F., Maasakkers, J. D., Adon, M., and Martin, R. V.: Worldwide biogenic soil NO_x emissions inferred from OMI NO2 observations, *Atmospheric Chemistry and Physics*, 14, 10 363–10 381, https://doi.org/10.5194/acp-14-10363-2014, 2014.

Wenger, A., Pugsley, K., O'Doherty, S., Rigby, M., Manning, A. J., Lunt, M. F., and White, E. D.: Atmospheric radiocarbon measurements to quantify CO2 emissions in the UK from 2014 to 2015, *Atmospheric Chemistry and Physics*, 19, 14 057–14 070, 2019.

Wu, D., Laughner, J. L., Liu, J., Palmer, P. I., Lin, J. C., and Wennberg, P. O.: A simplified non-linear chemistry transport model for analyzing NO2 column observations: STILT–NOx, *Geoscientific Model Development*, **16**, 6161–6185, https://doi.org/10.5194/gmd-16-6161-2023, 2023.

Zhao, C. and Wang, Y.: Assimilated inversion of NO_x emissions over east Asia using OMI NO2 column measurements, *Geophysical Research Letters*, 36, L06805, https://doi.org/10.1029/2008GL037123, 2009.

Appendix A

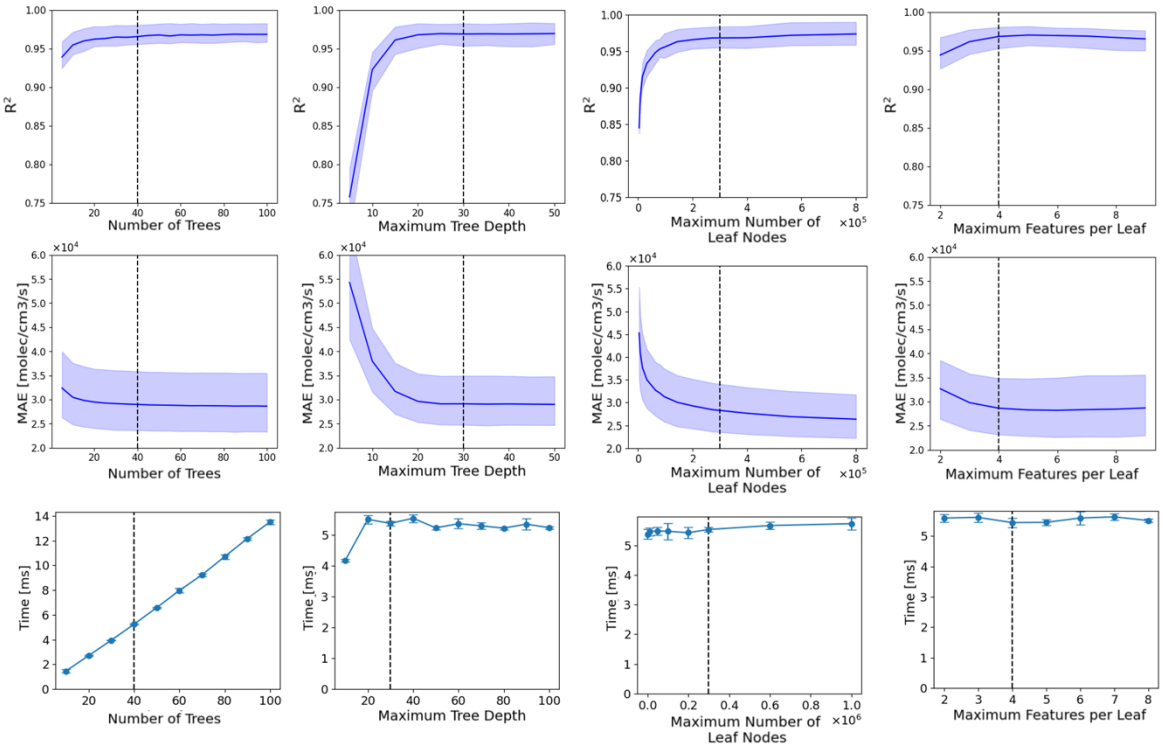


Figure A1: Impact of hyperparameter changes on random forest regression model performance for predicting NO_x chemistry rates. Plots show the effect of varying the number of trees, maximum tree depth, maximum leaf nodes, and maximum features per decision on mean R^2 MAE, and prediction time (shaded regions represent performance ranges across monthly models). Increased algorithm complexity improves R^2 and reduces MAE but increases prediction time. Optimal hyperparameters—40 trees, depth of 30, 300,000 leaf nodes, and 4 features per decision—achieve balanced performance with a prediction time of ~6 ms.

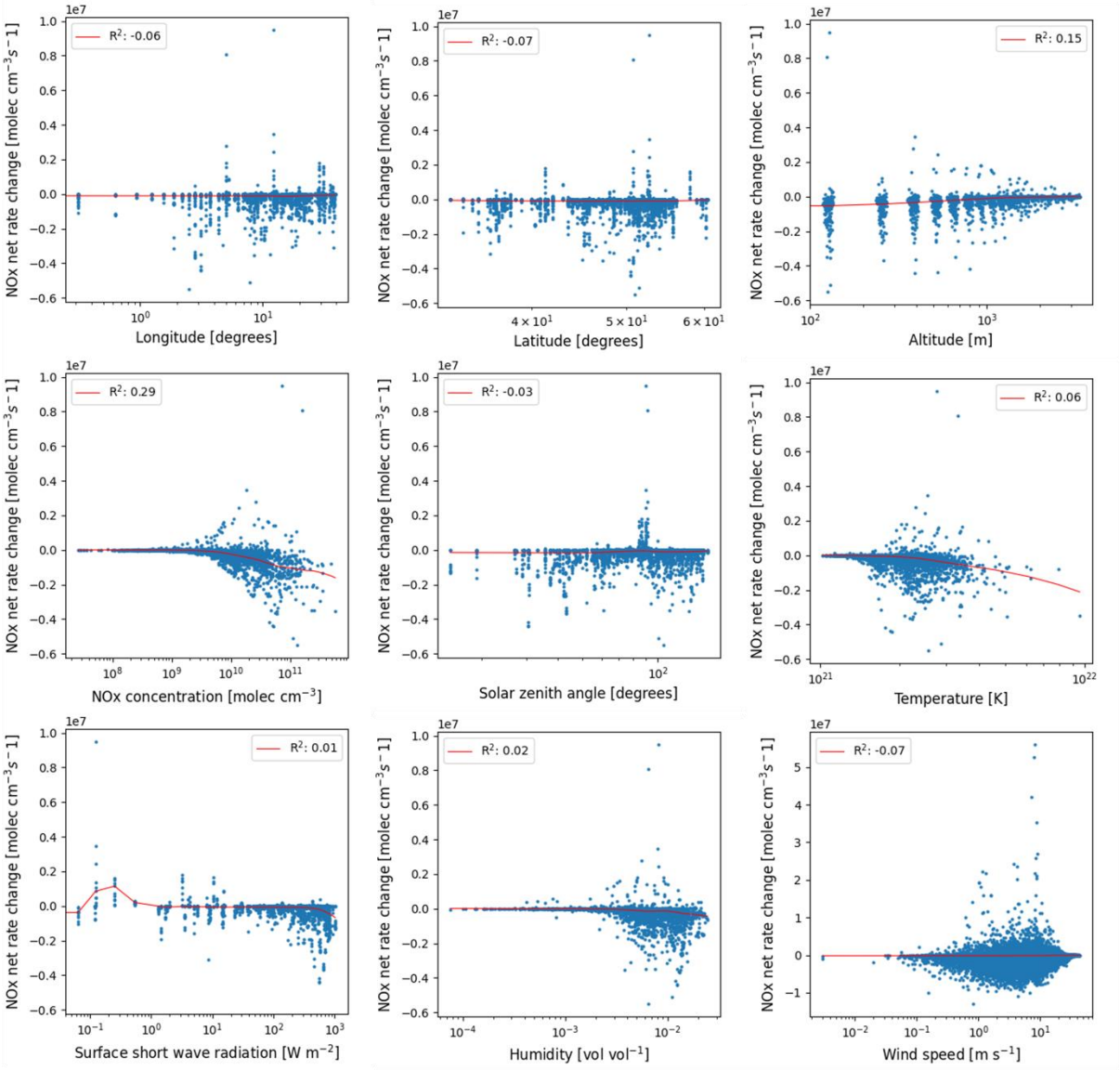
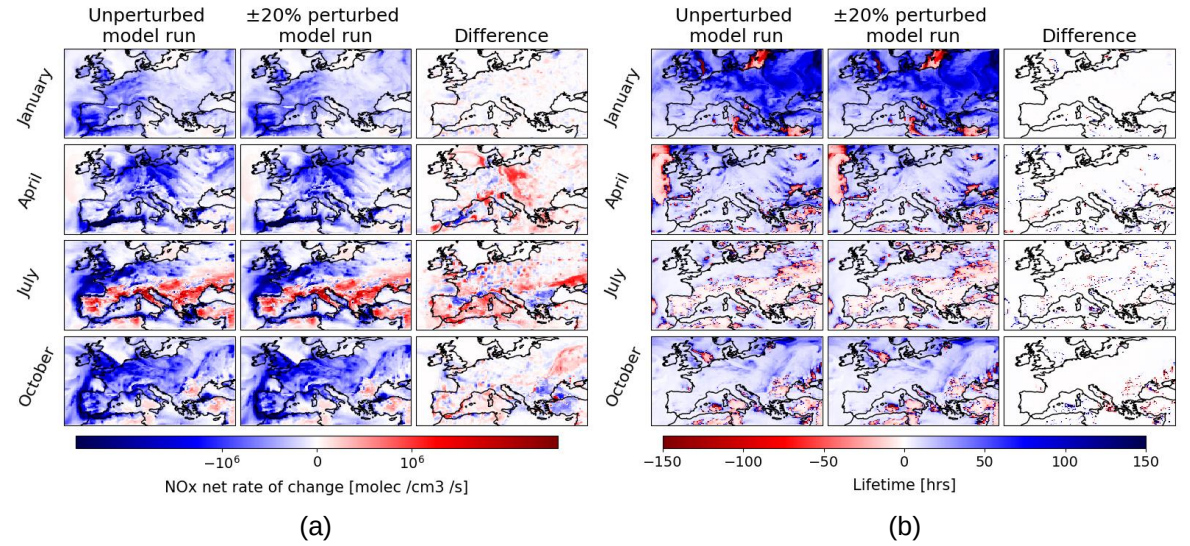
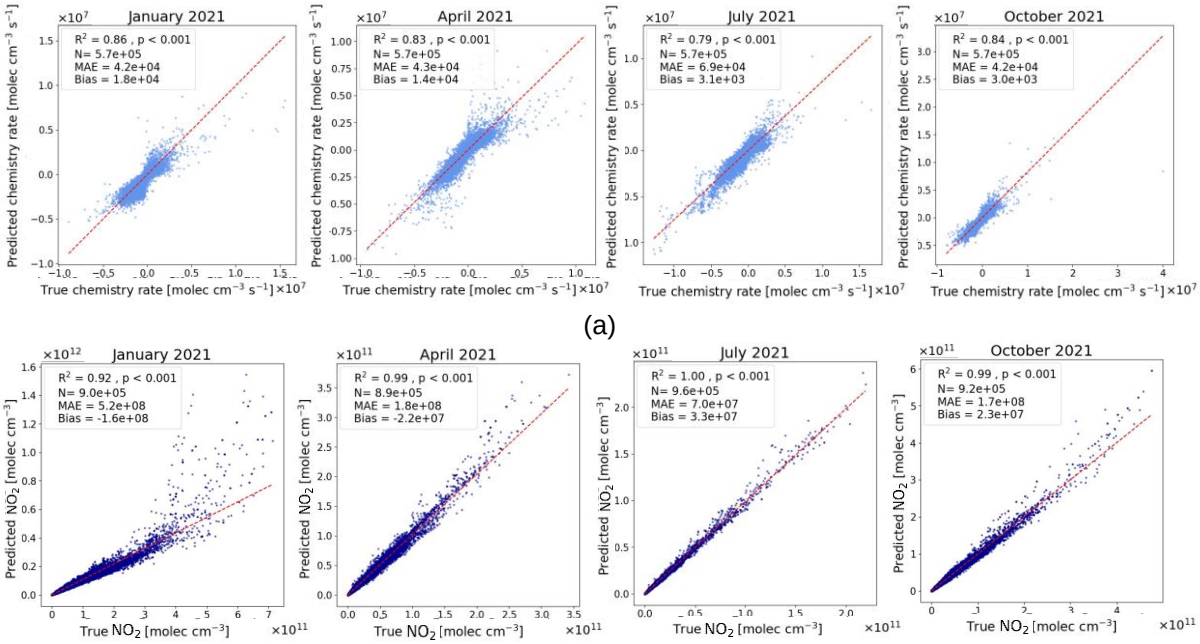


Figure A2: Individual relationships between the nine regression input parameters and the NO_x net rate of change. A LOWESS fit (red line) illustrates smoothed trends in the data, with R^2 values reported for each fit. Among the parameters, NO_x concentration, altitude, and temperature exhibit noticeable trends with chemistry rates, while the remaining parameters show little to no clear trends individually.



(a) (b)

Figure A3: The spatial distribution of the impact of ±20% emission perturbations on (a) the NO_x net rate of change, and (b) the atmospheric lifetime of NO_x. Overall, it is clear that the impact on the atmospheric lifetime is much smaller, due to its independence from the NO_x species concentration. Note that a negative lifetime of NO_x arises in areas where we have a net chemical production of NO_x.



(a)

(b)

Figure A4: Testing the regression models on 2021. (a) The random forest regression model for predicting the NO_x chemistry rate, (b) The reconstruction of NO_2 from NO_x using the random forest regression model for predicting the NO_2:NO ratio.

Appendix B: Comparison with TROPOMI

The NO_2 columns modelled by GEOS-Chem was compared directly with the TROPOMI data for assessment of agreement. Scatter plots between the two are shown in Fig. B1, where we found significant Pearson correlations ($p<0.001$) in all months. In January we observe a general positive bias, where the model is overestimating NO_2, while in July and October, a negative bias is seen.

The spatial distribution of the deviation between GEOS-Chem and TROPOMI is shown in Fig. B2. While there are clear areas of difference, it is notable that the general regions where we observe elevated levels of NO_2 are in alignment. In general, the spatial distribution of high-emission regions throughout Europe is fairly well understood. However, there is likely some error on the magnitudes of the emissions in the inventories used. This is likely to explain the majority of the areas of large bias between the model and the observations. However, it must be noted that other sources of error are present, which include model errors in transport processes, potential inaccuracies in the model meteorology used, errors in parameterising deposition processes, and the limiting factor of the model spatial resolution. Furthermore, there is also error on the TROPOMI measurements (largely characterised by the TROPOMI column precision value) including from instrument noise, cloud and aerosol interference, and vertical profile and sensitivity assumptions. Looking to Fig. 7, it is clear that there are many regions where the error between the model and observations is significantly smaller than the satellite precision, and for such areas the contribution of NO_x emissions is likely to be accurate.

On the whole, it is promising to the performance of the model that there is a general correlation of agreement between the model and satellite data. However, there is room for improvement in model agreement, and model inversions would be one approach to achieve this.

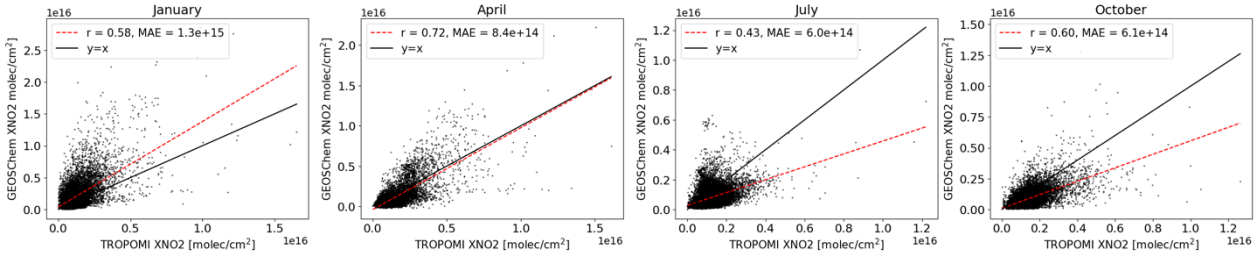


Figure B1: Correlation between modelled GEOS-Chem NO_2 columns and observed TROPOMI NO_2 for the four months of interest. The Pearson rank and mean absolute area are shown in the legend. The best-fit line (red-dashed) can be compared to the y=x line (black).

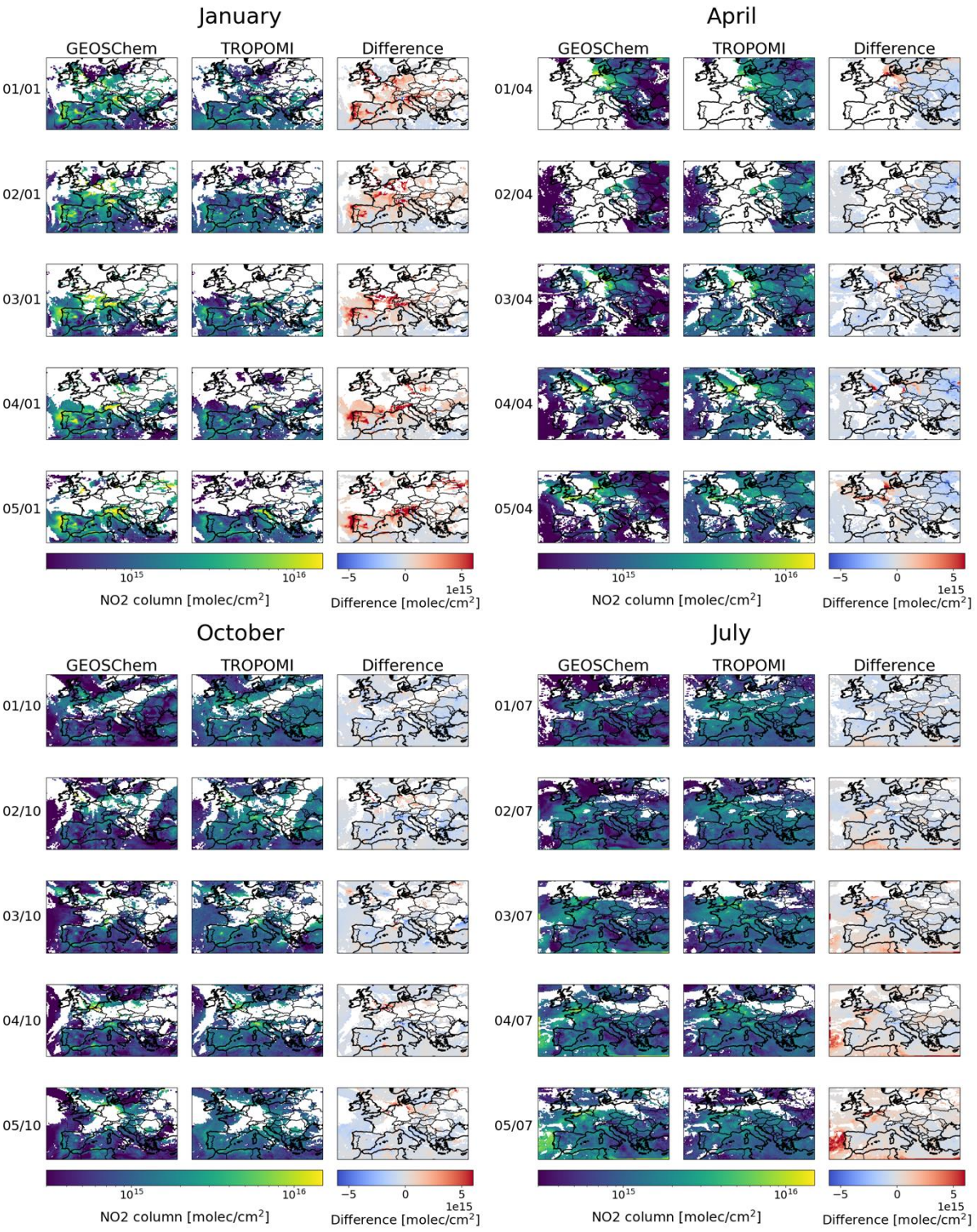


Figure B2: Comparison between GEOS-Chem and TROPOMI for 5 days in January, April, July, and October.

Document History

Version	Author(s)	Date	Changes
1.0	C. Schooling	November 28, 2024	
2.0	C. Schooling	November 29, 2024	Updated figure quality
3.0	A. Visser	December 6, 2024	Added Section 4.4
4.0	C. Schooling	December 18, 2024	Applied reviewer feedback

Internal Review History

Internal Reviewers	Date	Comments
Kuhlmann, Gerrit EMPA, Ingrid Luijkx, WUR	December 2024	