S1 General weather and chemical conditions

Figure S1 shows the flight paths for the 10 profiles and 3 horizontal scans overlayed over the coinciding TROPOMI NO_2 columns. Fig. S2 shows the measured and modeled potential Temperature profiles for the 10 vertical profiles.

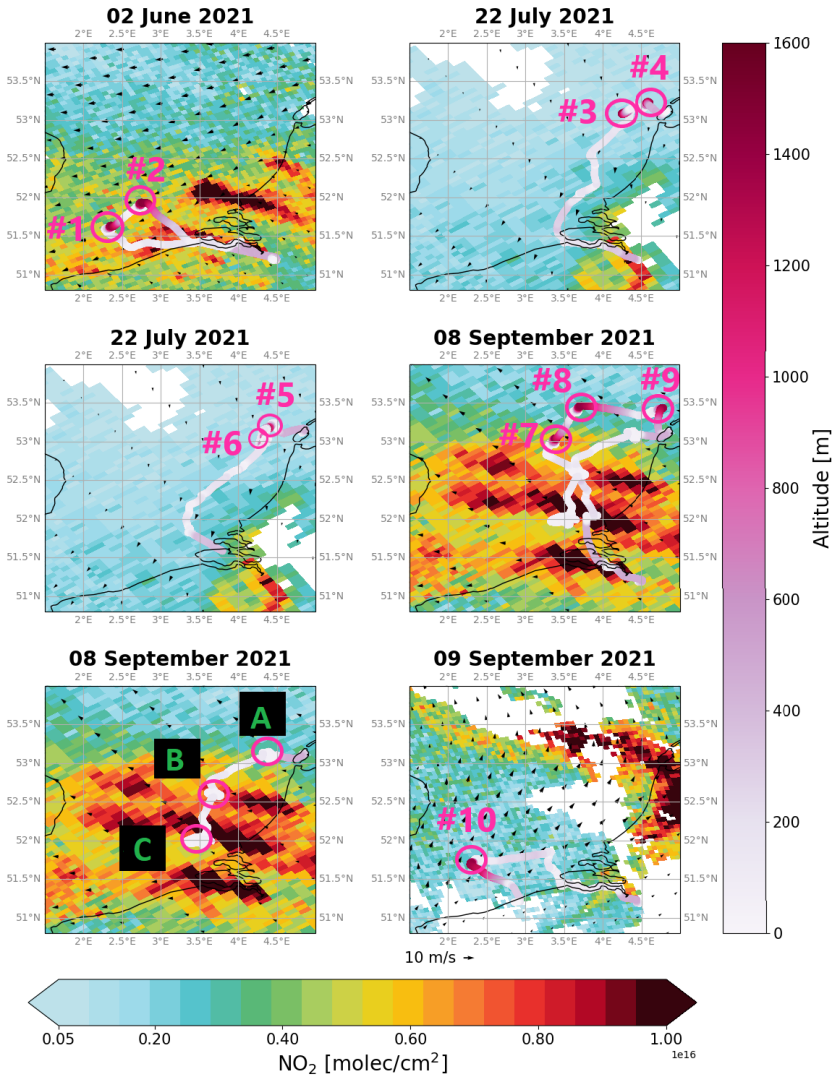
Below we summarize the general weather conditions relevant for interpreting the aircraft data during the four flight days.

2 June: Moderate winds from the east/southeast, over the North Sea around 10 m/s, wind gusts similar, and mostly cloud free. Warm weather, temperatures over the Netherlands >25°C. Air pollution from the Randstad is blown onto the southern North Sea with a east-southeast to west-north-west orientation.

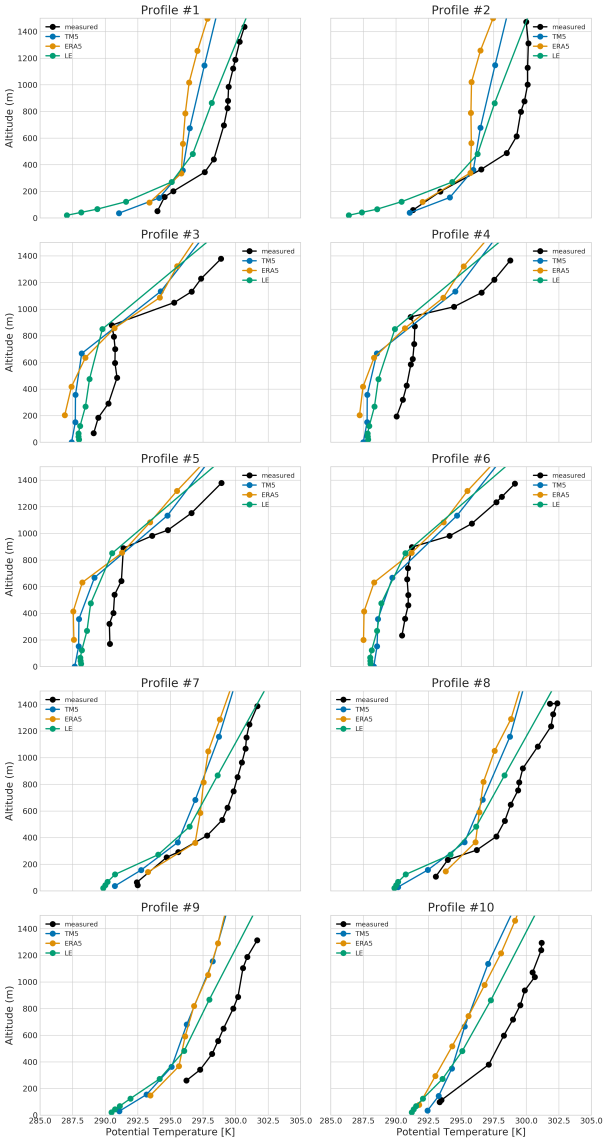
22/23 July: Light, mostly northerly winds mostly, with wind speeds of 3-7 m/s, wind gusts similar, mostly cloud free with some shallow cumulus and stratocumulus clouds over the North Sea. Pleasant weather, maximum temperatures over the Netherlands 20°-25°C. Chemical conditions are clean and northerly flow causes air pollution from the Netherlands to be blown away from the North Sea with a east-southeast to west-north-west orientation.

8 September: Light to moderate winds east/southeasterly, wind speeds 5-10 m/s, wind gusts similar. Nearly cloud free warm weather, maximum temperatures over the Netherlands >25°C. Air pollution from the Randstad is advected onto the southern North Sea.

9 September: Light to moderate winds from the south, wind speeds 5-10 m/s, wind gusts similar. Scattered low clouds and some stratocumulus fields, indicative of increasing atmospheric instability. Warm weather, maximum temperatures over the Netherlands >25°C. Air pollution from the south-western Netherlands is blown onto the southern North Sea with a north-south orientation. The Southernmost North Sea is relatively clean with a southwesterly flow.

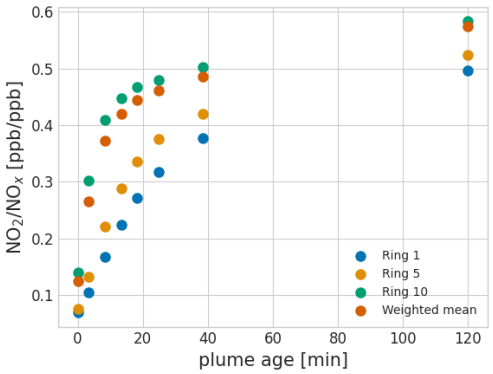


S 1: Overview of flight paths for ten vertical profiles (1-10) and three horizontal scans (A-C). Shown are the NO_2 column densities over the North Sea measured by TROPOMI at time of flight and the route of the aircraft as well as ECMWF 10 m wind indicated by the arrows.



S 2: Profiles of potential temperature of all flights as well as coinciding, TM5, CAMS ensemble mean and LE 2x2 km profiles.

S2 PARANOX



S 3: Relationship NO_2/NO_x ratio to binned plume age for PARANOX modeled plume for conditions during profile flight 1.

The PARANOX (PARametrization of Aircraft emitted NOX) model is a Gaussian plume model based on work from [Vila-Guerau de Arellano et al., 1990] which was successfully used for the parameterization of aircraft emissions and later for ship emissions [Vinken et al., 2011]. PARANOX simulates the dispersion of a plume in 10 concentric elliptical rings and the chemical evolution of the concentrations of several atmospheric trace gases inside it.
In this study we use the weighted average NO_2/NO_x ratio of the simulated aging plumes for neutral stability. For each of the 10 profiles a separate simulation was performed, using coinciding wind speeds from ERA5 reanalysis data, and background NO_x and O_3 values from the CAMS ensemble mean. For all simulations, the NOx emission strength was set at 60 g/s with 94% of the NO_x emitted as NO in line with [Eyring et al., 2005]. Fig. S3 shows the simulated NO_2/NO_x ratio for profile 1.

S3 Spatial representativeness of profile flights and other corrections

Profile number	MUMM flight	date	Post-hoc NO2 calibration [ppb]
#1	21082	02.06.2021	3.28
#2	21082	02.06.2021	4.44
#3	21116	22.07.2021	-0.04
#4	21116	22.07.2021	-0.02
#5	21117	22.07.2021	-0.80
#6	21117	22.07.2021	0.02
#7	21133	08.09.2021	-3.96
#8	21133	08.09.2021	-3.28
#9	21133	08.09.2021	-3.60
#10	21135	09.09.2021	-2.68

1. **Calibration.** We apply a post-hoc calibration of NO_2 mixing ratio C to fit the CAMS ensemble mean in the clean troposphere (above 250 m). The corrected volume mixing ratio C_{mumm} is the sum of the measured mixing ratio C_{meas} and the post-hoc calibration b: $C'_{mumm} = C_{meas} + b$ with b such that $\overline{X\,(C_{meas}+b)_{>250m}} = \overline{X\,(C_{CAMS})_{>250m}}$ where X is the mass density. The table above shows our post-hoc calibration values b for every flight. The biases are largely consistent during each day, consistent with the daily calibration routine executed prior to flight. Given the spread of calibrations during individual days, we assume a uncertainty of the bias correction of 0.5 ppb (mean of standard deviations for days 02.06., 22.07., 08.09.)).

2. **In-plume bias.** We expect any given area (e.g. a TROPOMI pixel) to be covered partially by a plume and partially by NO_2 background values. The coastguard flights actively approached ships and their plumes (see Sect. 2.2 in the main text) in order to measure the composition of the exhaust. Our measurements are therefore not necessarily representative of the mean concentrations over the entire area. Therefore, we use the representative random sampling of the transect flights to calculate an expected fraction of time in the plume for the lowest part of each profile (if it would have been a random sampling) and correct the observed measurements to fit this.
   To tackle this, we use 20-sec interpolated AIS ship location and the wind speed measured by the aircraft to calculate plume locations. We define an area of interest (AOI) which spans a rectangle from the minimal to maximal longitude/latitude. We use plume locations based on the AIS information on ship position and speed for all ships sailing within a distance of 0.5° margin on both latitude and longitude within 2 hours before the end of an aircraft profile. We use the mean wind direction and wind speed measured by the aircraft below 100 m to predict the presence of the plumes from any ship within the vicinity. We apply the projection method that was introduced by [Georgoulias et al., 2020]: the location of a plume at time t is simply the projected ship location from time t_0 multiplied with the wind speed in the direction of the wind: $x_{new} = x_{AIS} + u \cdot (t\text{-}t_0)$. We then count the number N of expected 20-second plume locations in the AOI and divide it by the size of the AOI A to calculate a so-called in-plume fraction $f_{AOI} = N_{AOI}/A_{AOI}$.

Taking the three transect flights (where the whole horizontal plane was sampled and therefore have a representative image of the area), we look at the NO_x measurements and define a measurement to be 'in plume' if the mixing ratio is at least 8 ppb above a background given by the 25 percentile of all NO_x measurements in said transects. Similar values for the threshold have been tested and yield similar results. From that we calculate the fraction of in plume measurements r as: $r_{AOI} = \frac{N_{inplume}}{N_{total}}$. Next, we assume a linear relationship between f and r and perform a linear fit through the three transect flights, yielding $r = c_1 * f + c_0$. This is our in-plume bias correction fit.

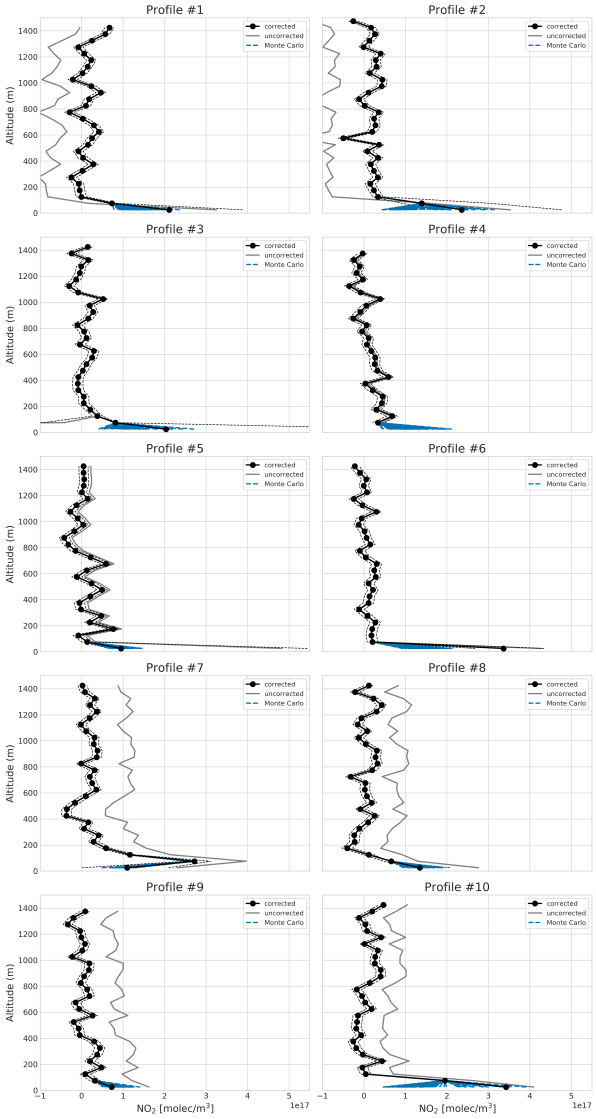
Using c_0 and c_1 from this fit, we calculate the expected r_{exp} for the lowest 100 m of each profile based on f_{AOI}. Additionally, we calculate the actual r_{actual} from the measured data. To correct each profile with the in-plume bias correction fit obtained from the transects we then calculate corrected mixing ratios $C''_{mumm,inplume} = C'_{mumm,inplume} * \frac{r_{exp}}{r_{actualy}}$ for in-plume measurements and $C''_{mumm,noplume} = C'_{mumm,noplume} * \frac{1-r_{exp}}{1-r_{actualy}}$ for measurements out of the plume. Since the linear fit is based on three data points only, we assume an uncertainty of 20% on r_{exp}.

To test the reliability, we additionally use a Monte Carlo simulation to calculate the NO_2 concentration in the lowest 50 m. Here, we first calculate the expected ratio of in/out of plume measurements r_{exp} as above and then draw random values of in/out of plume measurements of all profile flights, using the same in/out of plume definition as above. For every profile, we run 100 such simulations, using 20 measurements each (of which $20*r_{exp}$ are drawn from the 'in plume' pool). The 100 results give a range of expected values for the lowest 50 m and are largely in line with the first approach. The Monte Carlo approach gives a more multi-flight representation of the in- and out-of-plume measurements and might thus overestimate or underestimate NO_2 concentrations in plumes of weak or strong emitting ships, respectively.

3. **NO_2/NO_x ratio correction.** The NO_2 values are not directly measured but taken from the difference between NO_x and NO. If the two are not measured completely simultaneously but slightly shifted, a (strong) gradient in NO_x concentrations will lead to strongly negative NO_2 if the aircraft is measuring NO while in the plume, and NO_x when outside the plume. Similarly, when NO_x is measured in-plume and NO out-of-plume, the NO_2 values will be overestimated in respect to the measurement interval. To avoid such extreme NO_2 values, we run the Gaussian plume model PARANOX (see Supplemente S2) with NO_x and O_3 background levels similar to those during measurements. The modeled NO_2/NO_x ratios for rings 1 (center of plume), 5, 10 (edge of plume) and a volume weighted average for conditions during profile 1 are shown in Fig. S3.

   All NO_2 measurements in-plume with negative NO_2 or NO_2/NO_x ratios above 0.8 are then corrected to fit the weighted modeled ratio according to the plumes age. We assume an uncertainty for the modeled NO_2/NO_x ratio of 0.1.

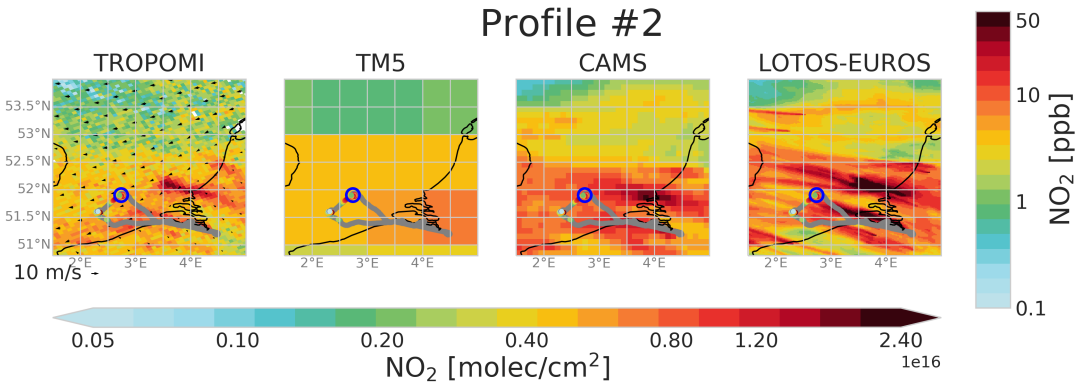
Fig. S4 shows the uncorrected and corrected profiles including the uncertainty estimate for the corrected data as well as the Monte Carlo simulations for the lowest level. Generally, the calibration bias dominates as can be seen from the offset between the gray and black bold lines. Below 150 m, the other corrections come into play, making the surface pollution more or less pronounced.



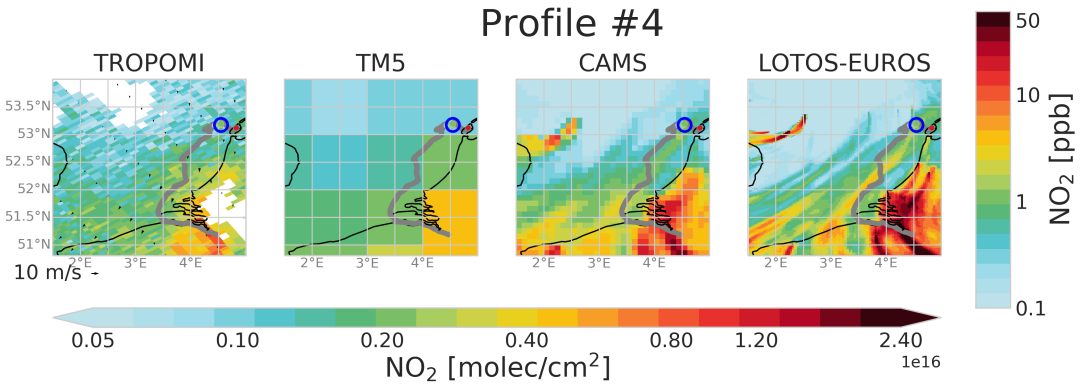
S 4: Aircraft measured profiles. The bold gray line shows the raw measured data, the bold black line the corrected profile, the dashed black lines the uncertainty estimate and the blue lines show the Monte Carlo simulation for the lowest 50 m.

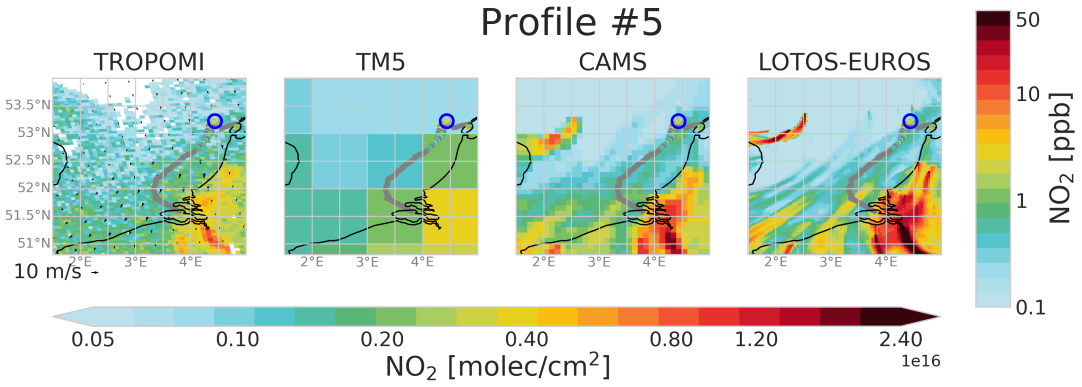
S4 Additional NO_2 maps

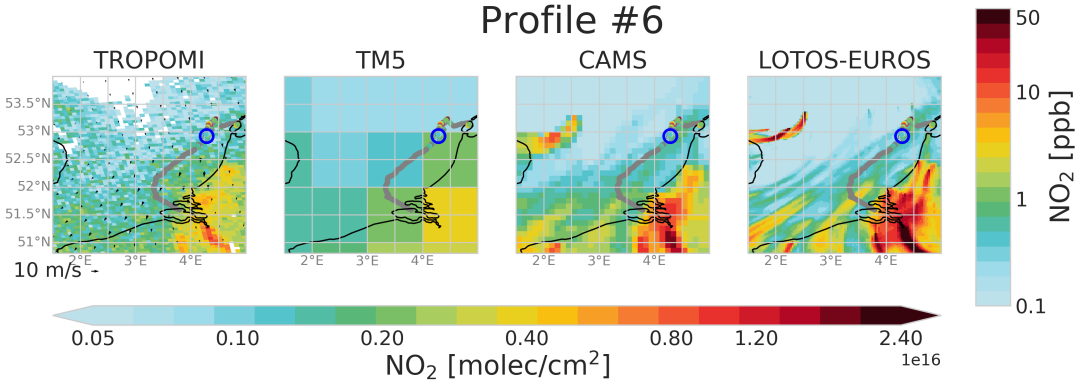
Fig. S5 shows the TROPOMI and modeled NO_2 column during the profile flights not presented in the main manuscript as well as the flight path and aircraft NO_2 measurements below 200 m.

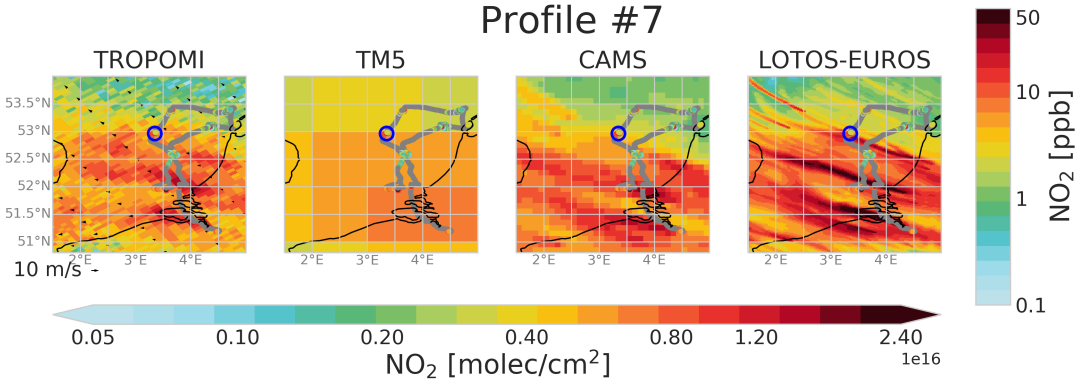


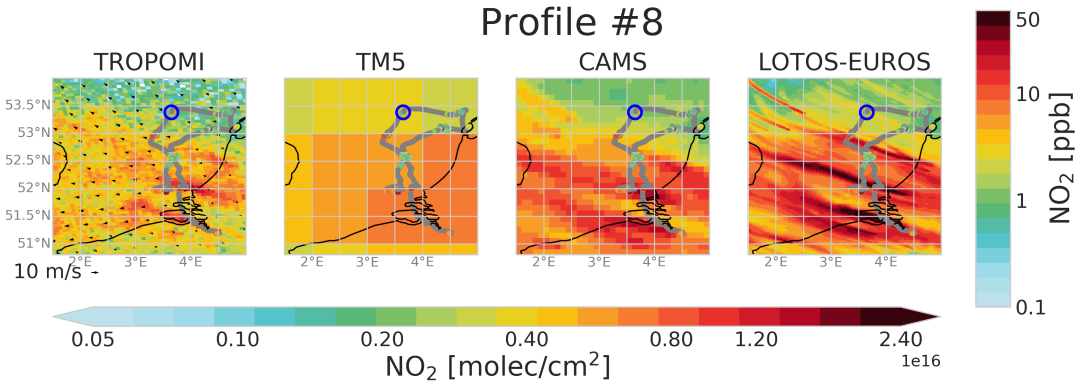
S 5: NO_2 columns as seen by TROPOMI and several model products for the time of the second profile measurement as indicated by the bottom colorbar. Overlayed are the aircraft measurements in grey for flights above 200 m and in colors below as indicated by the colorbar on the right as well as wind speed and direction by the arrows in the left panel.

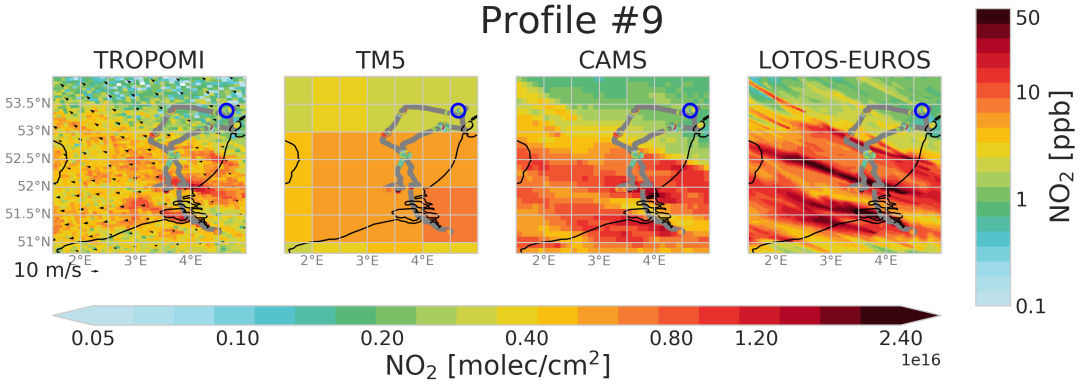


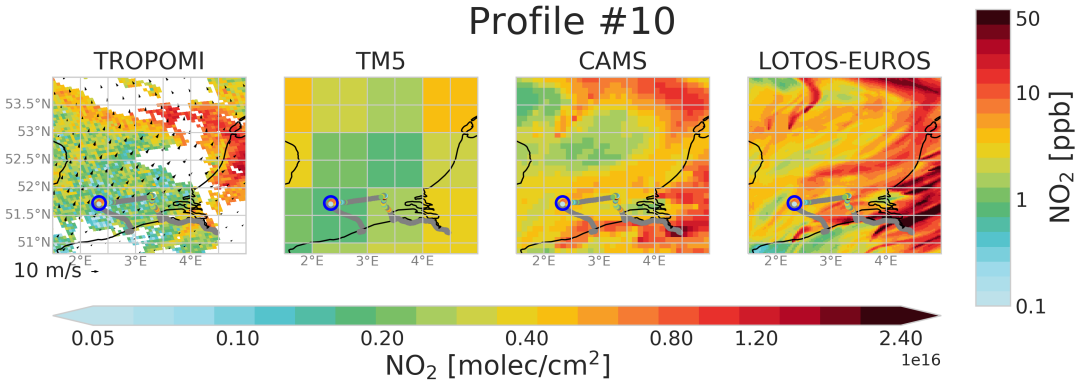

Profile #5
TROPOMI
TM5
CAMS
LOTOS-EUROS
53.5°N
53°N
52.5°N
52°N
51.5°N
51°N
2°E
3°E
4°E
10 m/s
0.05
0.10
0.20
0.40
0.80
1.20
2.40
1e16
NO2 [molec/cm2]
50
10
1
0.1
NO2 [ppb]


Profile #6
TROPOMI
TM5
CAMS
LOTOS-EUROS
53.5°N
53°N
52.5°N
52°N
51.5°N
51°N
2°E
3°E
4°E
10 m/s
0.05
0.10
0.20
0.40
0.80
1.20
2.40
1e16
NO2 [molec/cm2]
50
10
1
0.1
NO2 [ppb]

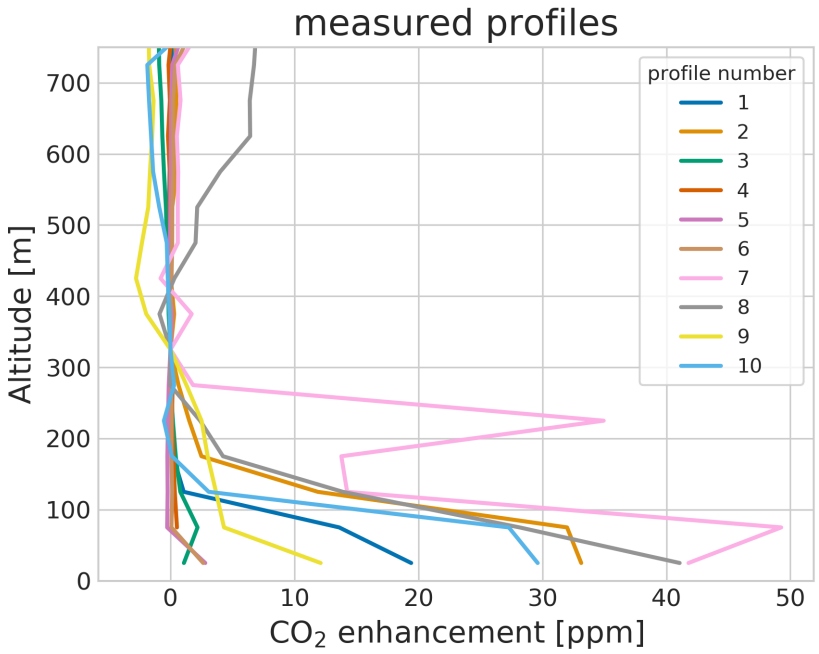

Profile #7
TROPOMI
TM5
CAMS
LOTOS-EUROS
53.5°N
53°N
52.5°N
52°N
51.5°N
51°N
2°E
3°E
4°E
10 m/s
0.05
0.10
0.20
0.40
0.80
1.20
2.40
1e16
NO2 [molec/cm2]
50
10
1
0.1
NO2 [ppb]


Profile #8
TROPOMI
TM5
CAMS
LOTOS-EUROS
53.5°N
53°N
52.5°N
52°N
51.5°N
51°N
2°E
3°E
4°E
10 m/s
0.05
0.10
0.20
0.40
0.80
1.20
2.40
1e16
NO2 [molec/cm2]
50
10
1
0.1
NO2 [ppb]


Profile #9
TROPOMI
TM5
CAMS
LOTOS-EUROS
53.5°N
53°N
52.5°N
52°N
51.5°N
51°N
2°E
3°E
4°E
10 m/s
0.05
0.10
0.20
0.40
0.80
1.20
2.40
1e16
NO2 [molec/cm2]
50
10
1
0.1
NO2 [ppb]


Profile #10
TROPOMI
TM5
CAMS
LOTOS-EUROS
53.5°N
53°N
52.5°N
52°N
51.5°N
51°N
2°E
3°E
4°E
10 m/s
0.05
0.10
0.20
0.40
0.80
1.20
2.40
1e16
NO2 [molec/cm2]
50
10
1
0.1
NO2 [ppb]

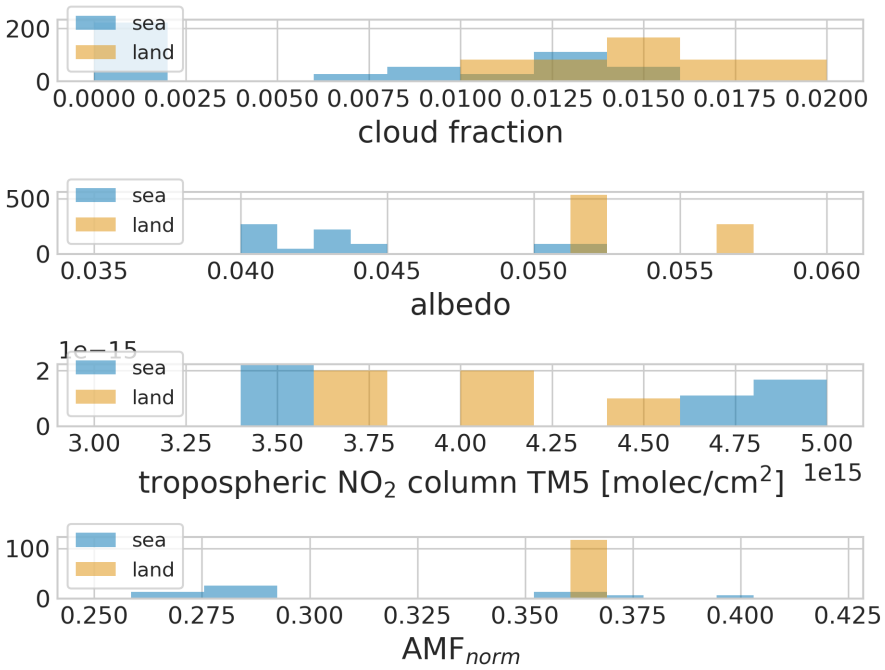
S5 CO_2 profiles



S 6: CO_2 enhancement profiles for the ten spiral flights. As the CO_2 sensor was not calibrated, the 300-350 m value is subtracted to visualize the enhancement in the mixed layer.

Fig. S6 shows the CO_2 enhancements in the mixed layers measured for the ten profiles. The observed mixing layer height of ¡ 150m agrees well with the observations in the NO_2 profiles. The CO_2 concentrations are measured directly with a temporal resolution of 1 Hz and no corrections are applied here.

S6 Sea-land contrast



S 7: Normalized distributions of cloud fraction, surface albedo, TM5 a priori columns and normalized tropospheric AMF over land (yellow) and sea (blue).

We compare tropospheric AMFs over sea at the time and location of the aircraft measurements to AMFs over land at the Cabauw tower in the same TROPOMI orbits. For this analysis, we normalize tropospheric AMFs for viewing and solar geometry by dividing the topospheric AMFs provided in the TROPOMI files by the geometric AMFs. To make sure the observed differences are caused by the differences in surface properties, we include only scenes with similar a priori columns and cloudiness levels. Fig. S7 shows the distribution of cloud fractions, albedo, a priori columns and AMFs for the land and sea scenes. We find on average 20% lower tropospheric AMFs over the North Sea compared to land given similar overall retrieval conditions.

References

[Eyring et al., 2005] Eyring, V., Köhler, H. W., Lauer, A., and Lemper, B. (2005). Emissions from international shipping: 2. Impact of future technologies on scenarios until 2050. *Journal of Geophysical Research D: Atmospheres*, 110(17):183–200.

[Georgoulias et al., 2020] Georgoulias, A. K., Boersma, K. F., Van Vliet, J., Zhang, X., Van Der A, R., Zanis, P., and De Laat, J. (2020). Detection of NO2 pollution plumes from individual ships with the TROPOMI/S5P satellite sensor. *Environmental Research Letters*, 15(12):124037.

[Vila-Guerau de Arellano et al., 1990] Vila-Guerau de Arellano, J., Talmon, A., Part, P. B. A. E., and 1990, u. (1990). A chemically reactive plume model for the NO-NO2-O3 system. *Atmospheric Environment*, 24A(8):2237–2246.

[Vinken et al., 2011] Vinken, G. C., Boersma, K. F., Jacob, D. J., and Meijer, E. W. (2011). Accounting for non-linear chemistry of ship plumes in the GEOS-Chem global chemistry transport model. *Atmospheric Chemistry and Physics*, 11(22):11707–11722.