

# Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations



ESA Contract No. 4000135355/21/I-DT-Ir

## Deliverable D4.3: Validation Report

Date: 2026-05-12

Version: 1.3

**Contributing authors:**

Andreas Richter, Kai Krause (IUP-UB)

Martina M. Friedrich (IASB-BIRA)

Jos van Geffen, Ankie Pijters (KNMI)

## Table of Contents

|     |                          |    |
|-----|--------------------------|----|
| 1   | Introduction .....       | 4  |
| 2   | Validation method .....  | 4  |
| 3   | Validation results ..... | 7  |
| 3.1 | MAPA .....               | 11 |
| 3.2 | MMF .....                | 17 |
| 4   | Summary .....            | 21 |
|     | Appendix .....           | 23 |
|     | Scatterplots MAPA .....  | 23 |
|     | Scatterplots MMF .....   | 53 |

# 1 Introduction

The NO<sub>2</sub> field around a MAX-DOAS station can be inhomogeneous, especially for urban environments. Knowledge of the horizontal inhomogeneities in the NO<sub>2</sub> is useful and important for the interpretation of the data of a MAX-DOAS station, and especially for validation of satellite instruments.

The objective of FRM4DOAS WP-4 is to define a procedure for the characterization of the horizontal NO<sub>2</sub> field and the inhomogeneities within it for MAX-DOAS measurements in urban areas. The procedure needs to be applicable for all stations, robust for different situations, and fast enough to work in the FRM4DOAS processing chain.

To identify inhomogeneities, three different indicators were defined:

1. Azimuthal inhomogeneity
2. Temporal inhomogeneity
3. Combined azimuthal and temporal inhomogeneity

Each indicator was calculated as an absolute indicator in molec/cm<sup>2</sup> and a relative indicator, normalized by a corresponding median VCD. Additionally, representative averaged columns were defined and calculated from VCDs filtered for inhomogeneity based on these indicators. The derived parameters as well as the underlying algorithms are presented in the ATBD of WP4 (D-4.2) in detail.

# 2 Validation method

The inhomogeneity indicators were implemented and tested for several urban FRM4DOAS stations and multiple years of measurements were used for each station. This dataset was then matched to S5p tropospheric VCDs, and the influence of filtering using the inhomogeneity indicators on the correlation between both data sources was investigated. The goal was to determine whether satellite validation could be improved by incorporating information about the horizontal inhomogeneity of the NO<sub>2</sub> field.

To match both datasets, the time difference in measurements had to be lower than 10 minutes, with a distance between S5p pixel and station below 20 km and a S5p pixel qa\_value => 0.5. Additionally, the MAX-DOAS data has not only been matched to a single S5p pixel but also to 3x3 and 5x5 averages of S5p pixels around the matched pixel. For Athens, only S5p pixels located above the city itself are considered, as the instrument is located at the edge of the city and some azimuth direction only scan into rural areas, while other scan towards the harbour and more urban parts of the city. For Thessaloniki only azimuth direction from south to north-



west were used because hills and mountains only allow the MAX-DOAS instrument to scan in these directions.

Also, the four viewing directions of Mainz, which are currently delivered as four different instruments (Mainz\_1676\_1, Mainz\_1676\_2, Mainz\_1676\_3, Mainz\_1676\_4), were combined during the processing and will be summarized as Mainz\_1676 in the discussion of the results.

This initial dataset has then been further refined to use only S5p qa\_values equal to 1, MAX-DOAS cloud flag equal to 0, and MAX-DOAS qa\_flag smaller than 2. In this document, only this stricter dataset will be presented and used for all investigations. Comparison between S5p and MAX-DOAS using MAX-DOAS qa\_flag equal to 0 would have left only very few matching observations and was therefore not further considered for the analysis.

For the validation, linear regressions and orthogonal distance regressions were performed between the S5p tropospheric NO<sub>2</sub> VCDs and the respective MAX-DOAS VCDs for all stations and both profiling algorithms (MAPA and MMF). The regressions are carried out for unfiltered MAX-DOAS VCDs, MAX-DOAS VCDs that have been filtered by the inhomogeneity indicators, and the representative MAX-DOAS VCDs, which have been calculated as averages out of prefiltered data (more details in the ATBD). The filtered MAX-DOAS VCDs have been filtered by either their absolute inhomogeneity indicator, their relative inhomogeneity indicator, or the combination of both. When absolute and relative indicators are combined, the absolute indicator is used for VCDs lower than the station's median VCD and the relative indicator for VCDs higher than the station's median.

Figure 1 and Figure 2 show histograms of the inhomogeneity indicators in the MMF product for all stations. For each station, the histograms show positively skewed distributions. The results for MAPA are very similar and therefore not shown here. Based on the histograms, relative thresholds of 20% and absolute thresholds of  $1.0 \times 10^{15}$  molec/cm<sup>2</sup> were chosen as thresholds for inhomogeneity as a compromise that works reasonably well for all stations.

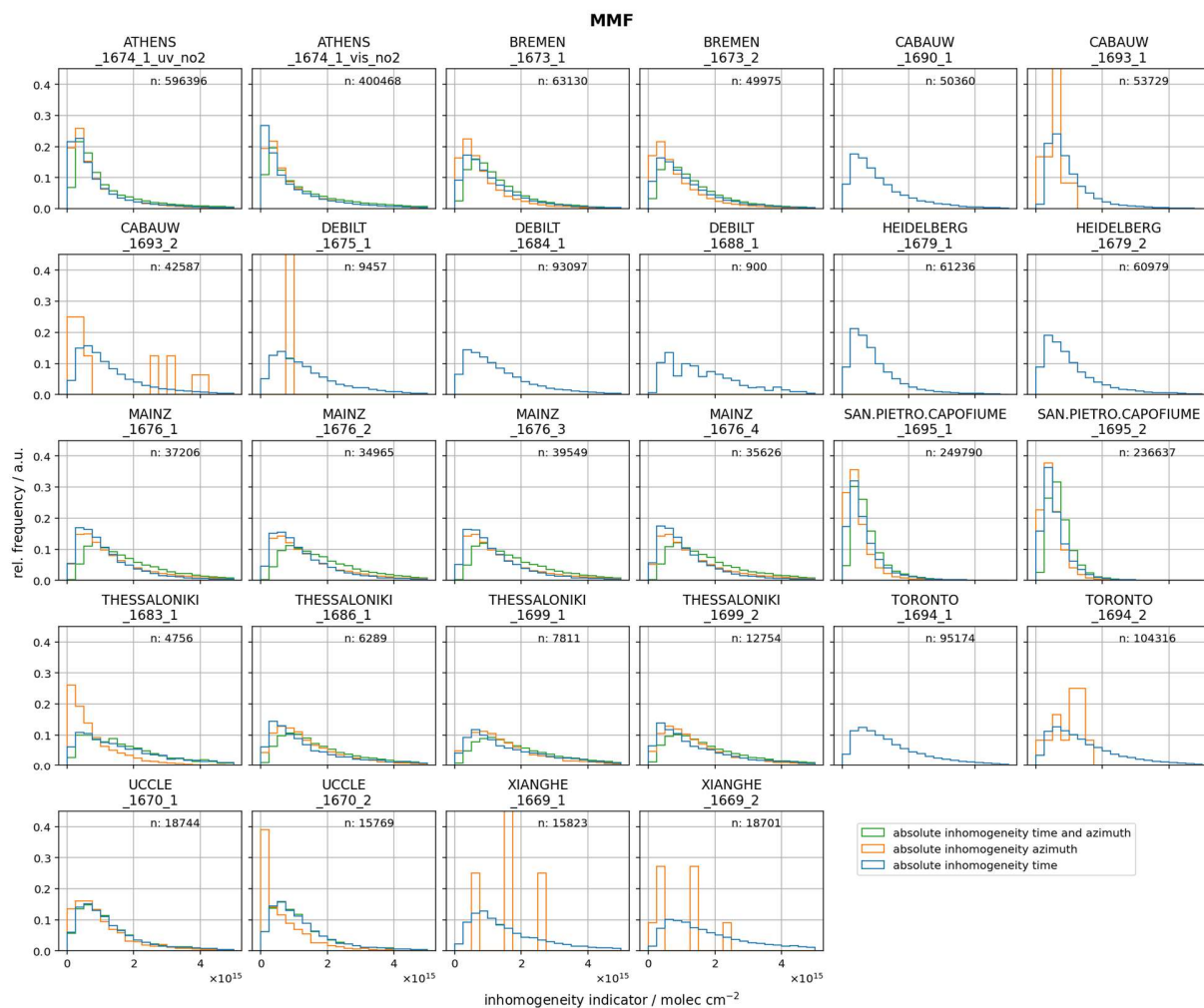


Figure 1: Histograms of the different absolute inhomogeneity indicators of the MMF product for each station.

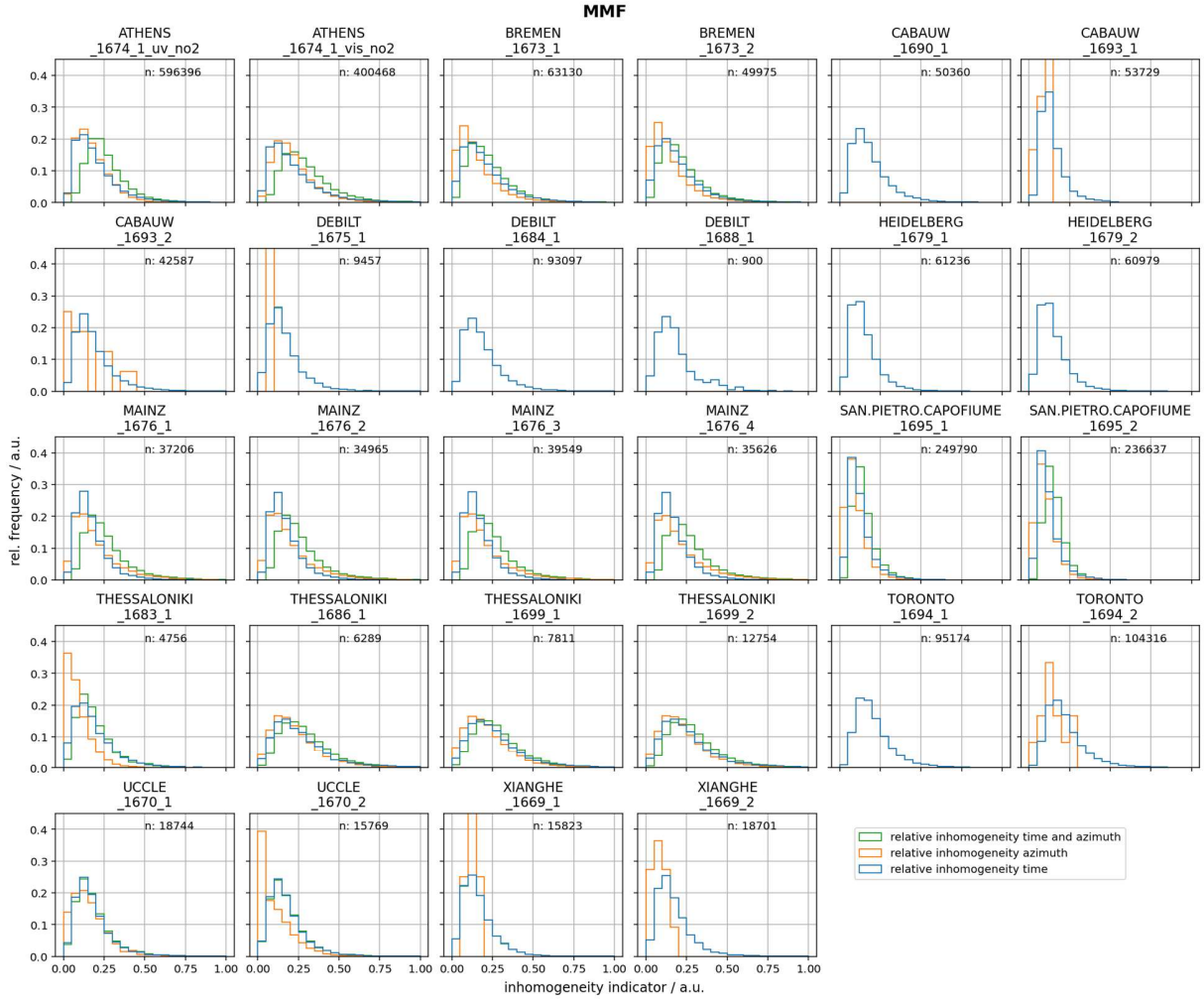


Figure 2: Histograms of the different relative inhomogeneity indicators of the MMF product for each station.

In all cases, regressions were performed with the single pixel S5p value, the 3x3 average, and the 5x5 average. Additionally, the influence of the distance on the regression results between the MAX-DOAS station and the S5p pixel has been investigated by carrying out the regressions a second time, but using matching VCDs only when the distance between the station and the S5p pixel was below 5 km.

### 3 Validation results

Figure 3 shows the results of the validation for the Bremen station using the MAPA VCDs in the ultraviolet wavelength range against the matched S5p tropospheric VCDs (single pixel), with a maximum distance between station and satellite pixel of 20 km. In the first three columns, the scatterplots and regression results for filtering using a specific inhomogeneity indicator are shown, starting with filtering for inhomogeneity in time, filtering for inhomogeneity in azimuth,

and filtering for the combined inhomogeneity in time and azimuth. The first plot in the last column shows the result for the unfiltered values, while the last plot in this column shows the results for the representative column calculated for VCDs that have been filtered for the combined azimuthal and temporal inhomogeneity. The combined setting uses the absolute value for VCDs below the station's median, while relative values are used for VCDs above the station's median. In all filtered cases, the correlation coefficients increase compared to the unfiltered dataset, the slopes get closer to 1, and the RMS decreases. Reducing the distance between S5p pixel and station further increases the correlation coefficients at (Figure 4), although the improvement is more pronounced when filtering for relative or the combination of absolute and relative inhomogeneity. The slopes do not improve compared to the dataset with a maximum distance of 20 km. Generally, for this station, the filtering results in a homogeneous removal of values in a wider range of VCDs and the correlations improve when the VCDs are filtered for inhomogeneity, and the best results are found when the representative column is used instead of the individual VCDs.

For this station the results using the MMF VCDs are very similar, filtering improves correlation, especially when filtering by a combination of relative and absolute indicators and the best regressions results are found when the representative column filtered for temporal and azimuthal inhomogeneity is used (see Figure 5). Similar plots for all stations and both retrieval algorithms can be found in the appendix.

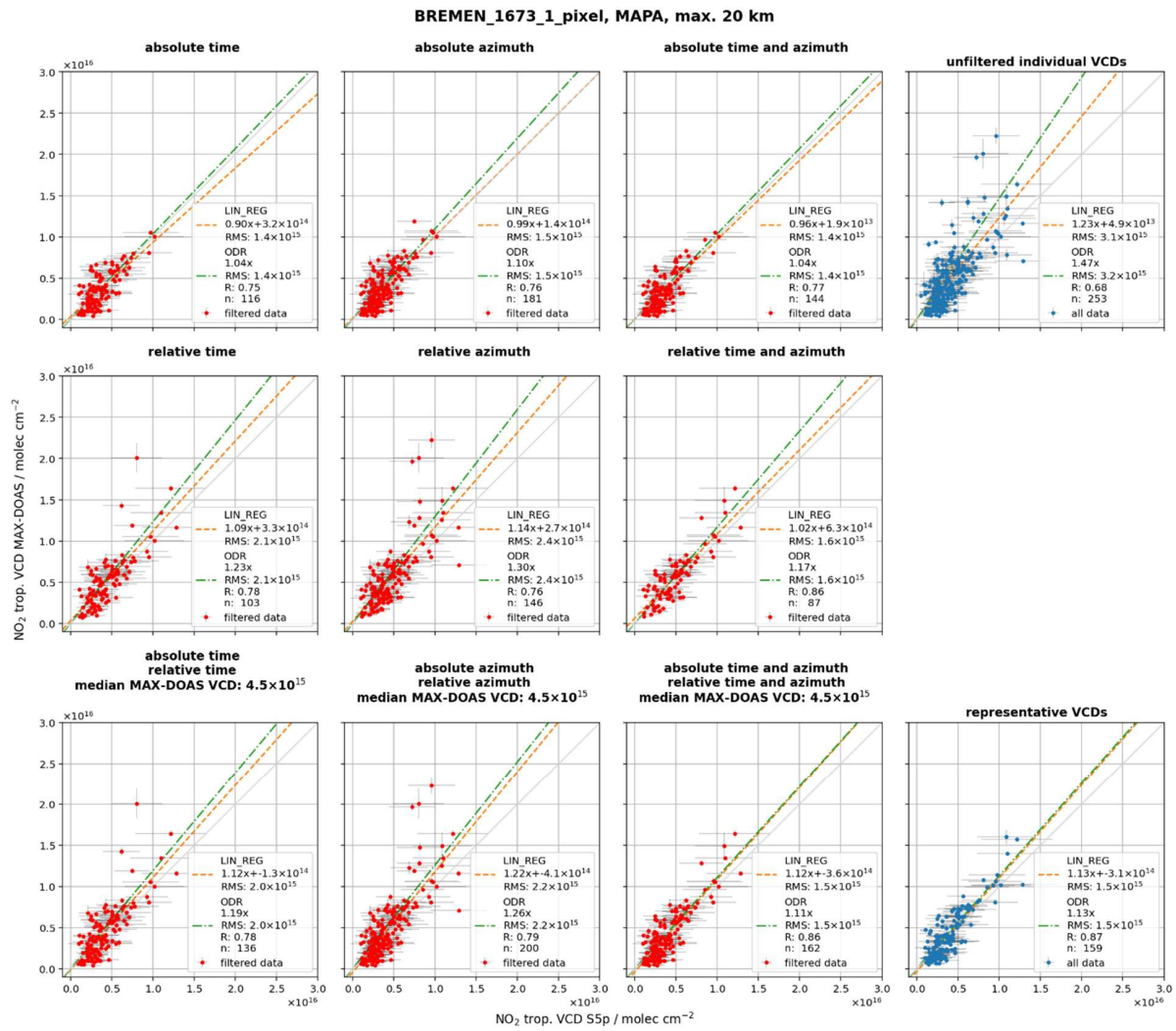


Figure 3: Scatterplots of S5p and MAPA VCDs for FRM4DOAS station Bremen\_1673\_1, with a maximum distance between station and assigned S5p pixel within a distance of 20 km. In the first three columns, the scatterplots and regression results for filtering using a specific inhomogeneity indicator are shown, starting with filtering for inhomogeneity in time, filtering for inhomogeneity in azimuth, and filtering for the combined inhomogeneity in time and azimuth. The first plot in the last column shows the result for the unfiltered values, while the last plot in the fourth column shows the results for the representative column calculated for VCDs that have been filtered for the combined azimuthal and temporal inhomogeneity. In the first row the VCDs in columns one to three are filtered by absolute inhomogeneity indicators, while in the second row relative inhomogeneity indicators are used, and finally in the third row a combination of relative and absolute indicators is used, where VCDs below the stations median were filtered using absolute indicators and VCDs above the stations median were filtered using relative indicators.

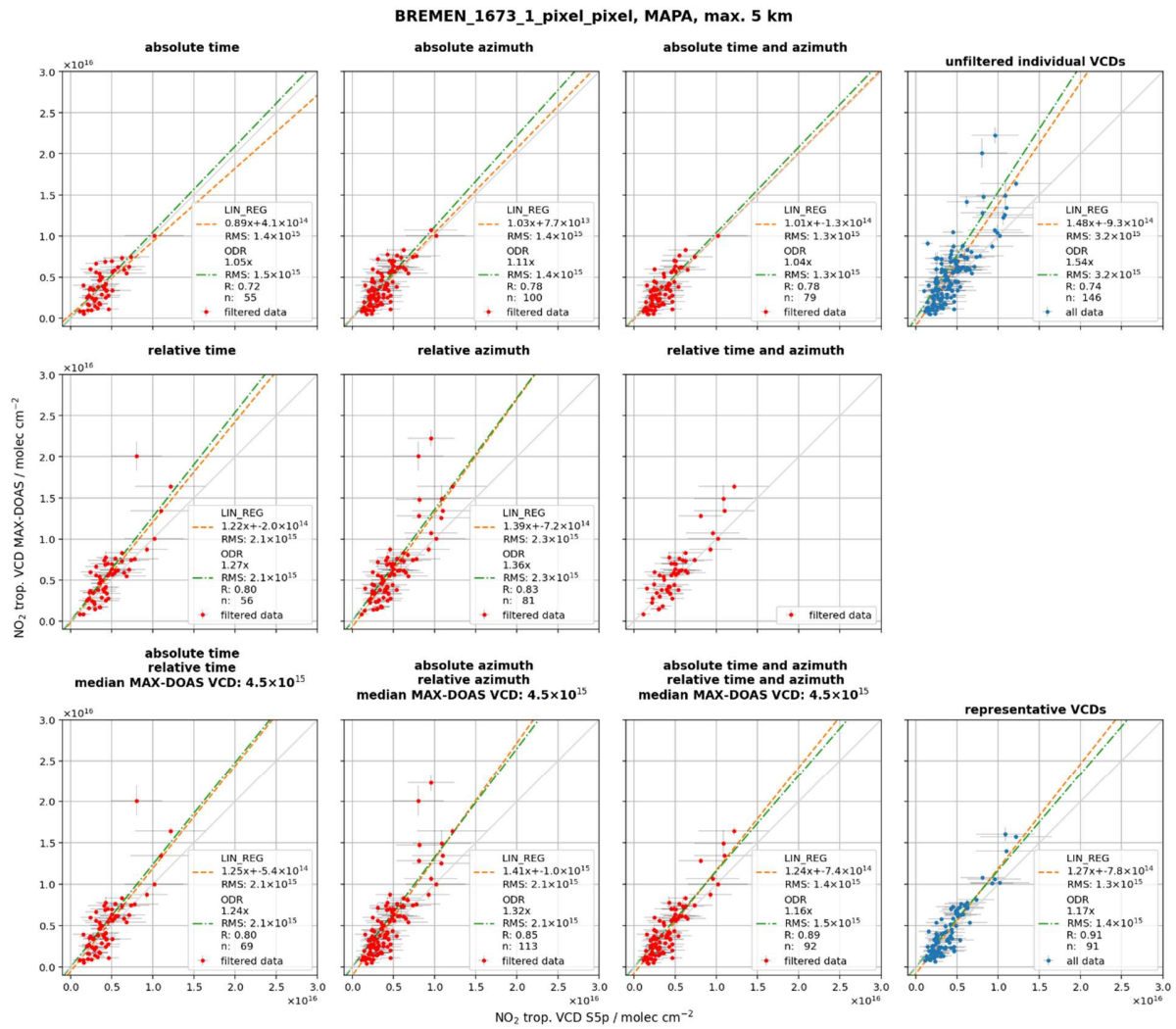


Figure 4: Scatterplots of S5p and MAPA VCDs for FRM4DOAS station Bremen\_1673\_1, with a maximum distance between station and assigned S5p pixel of 5 km.

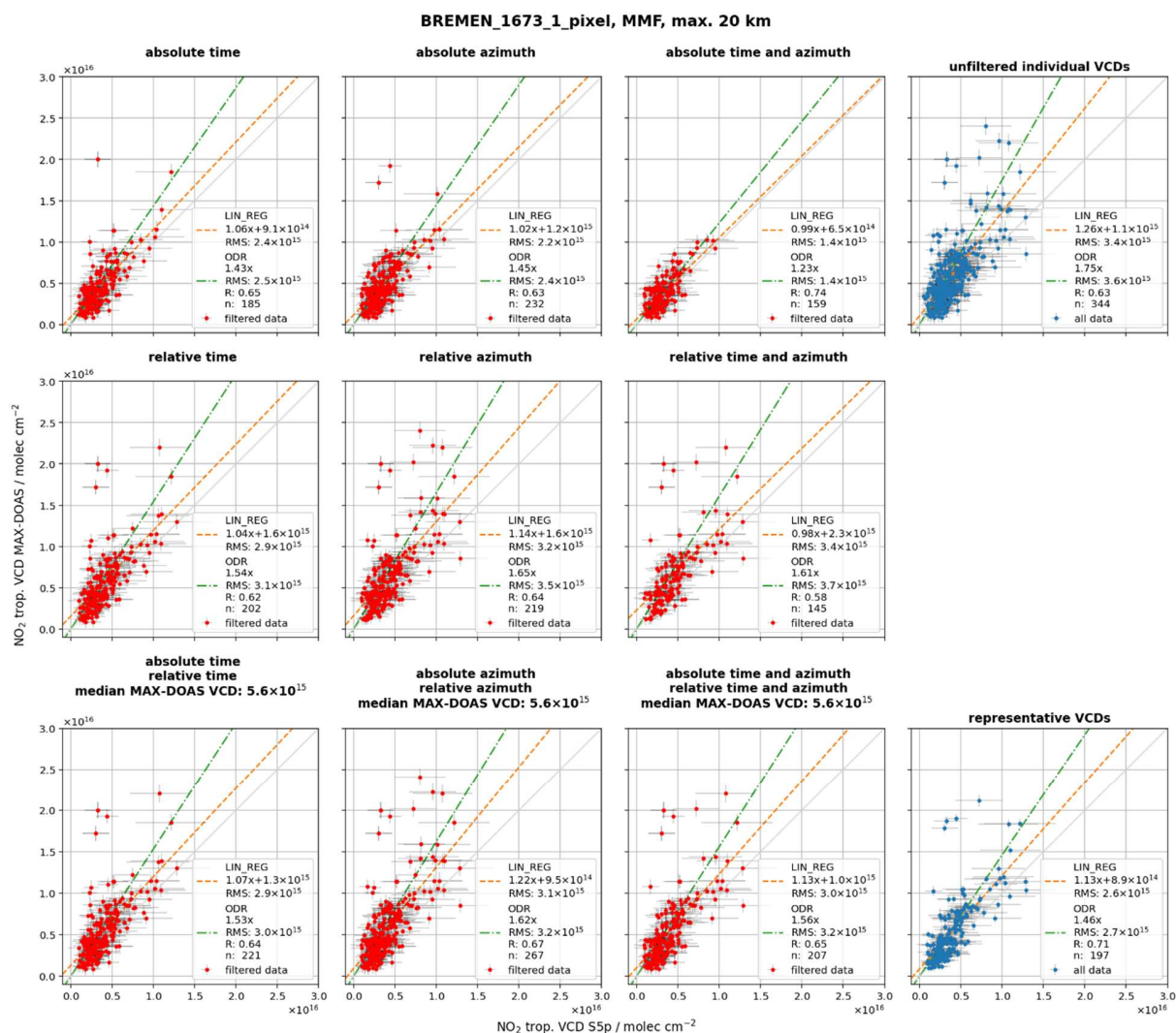


Figure 5: Same as Figure 3 but using MMF VCDs.

For a more detailed discussion of the results for each station, MMF and MAPA are presented separately in the following chapters.

### 3.1 MAPA

For comparison, Figure 6 shows the results for Athens in the visible wavelength range with a maximum distance between S5p and station of 20 km. Filtering in these cases mostly removes high VCDs and only leaves a cluster of low VCDs in the comparison to S5p. However, also here there is an improvement of slope, correlation coefficient, and RMS compared to the unfiltered datasets and the representative column shows the best results.



Filtering for individual inhomogeneity indicators does not always lead to better results, and especially thresholds for the absolute indicators need to be adapted to individual stations. Figure 7 shows the results for the Thessaloniki station (Thessaloniki\_1686\_1), and the chosen absolute threshold of  $1 \times 10^{15}$  molec/cm<sup>2</sup> clearly is not a good choice for this station, as only a cluster of low VCDs is left for comparison, and the total number of data points is reduced to below 50. In such cases, the regression results are not displayed. Nevertheless, the correlations improve when filtering for the combined temporal and azimuthal inhomogeneity using the combination of relative and absolute thresholds.

An overview of all orthogonal distance regression results for all stations using a maximum distance between pixel and station of 20 km is shown in Figure 8. This figure also shows the results for 3x3 and 5x5 pixel averaged S5p data in each column of the plot. The different filter settings are colour-coded and the representative column, filtered for combined inhomogeneity in time and azimuth is also shown. If there is no value shown for a specific station, there were fewer than 50 matched datapoints for the whole time period. In Mainz the azimuthal inhomogeneity filtered individual VCDs show similar results to the representative columns that have been filtered for the combined temporal and azimuthal inhomogeneity, the representative columns however still show the best correlations.

Generally, filtering improves the results for all stations, especially filtering using the combined temporal and azimuthal inhomogeneity indicators. In most cases, the representative columns show better results than the individually filtered and matched VCDs. The biggest improvements are found for Athens and Bremen. The improvement is also found for the averaged S5p pixel data, but has influence on the slopes. For Athens and Bremen, the slopes are closer to 1 for the 3x3 and 5x5 averages, while for San Pietro Capofiume, DeBilt, Uccle and Xianghe the slopes are similar to regression against individual S5p pixels.

The results for the dataset with a maximum distance of 5 km are similar, and the reduced distance mostly has an effect on the initial unfiltered correlations (see Figure 9). The results using linear regression are shown in Figure 10 and are similar to the ones for orthogonal distance regression, although for most stations the slopes here are closer to 1, with the exception of Heidelberg and Mainz.



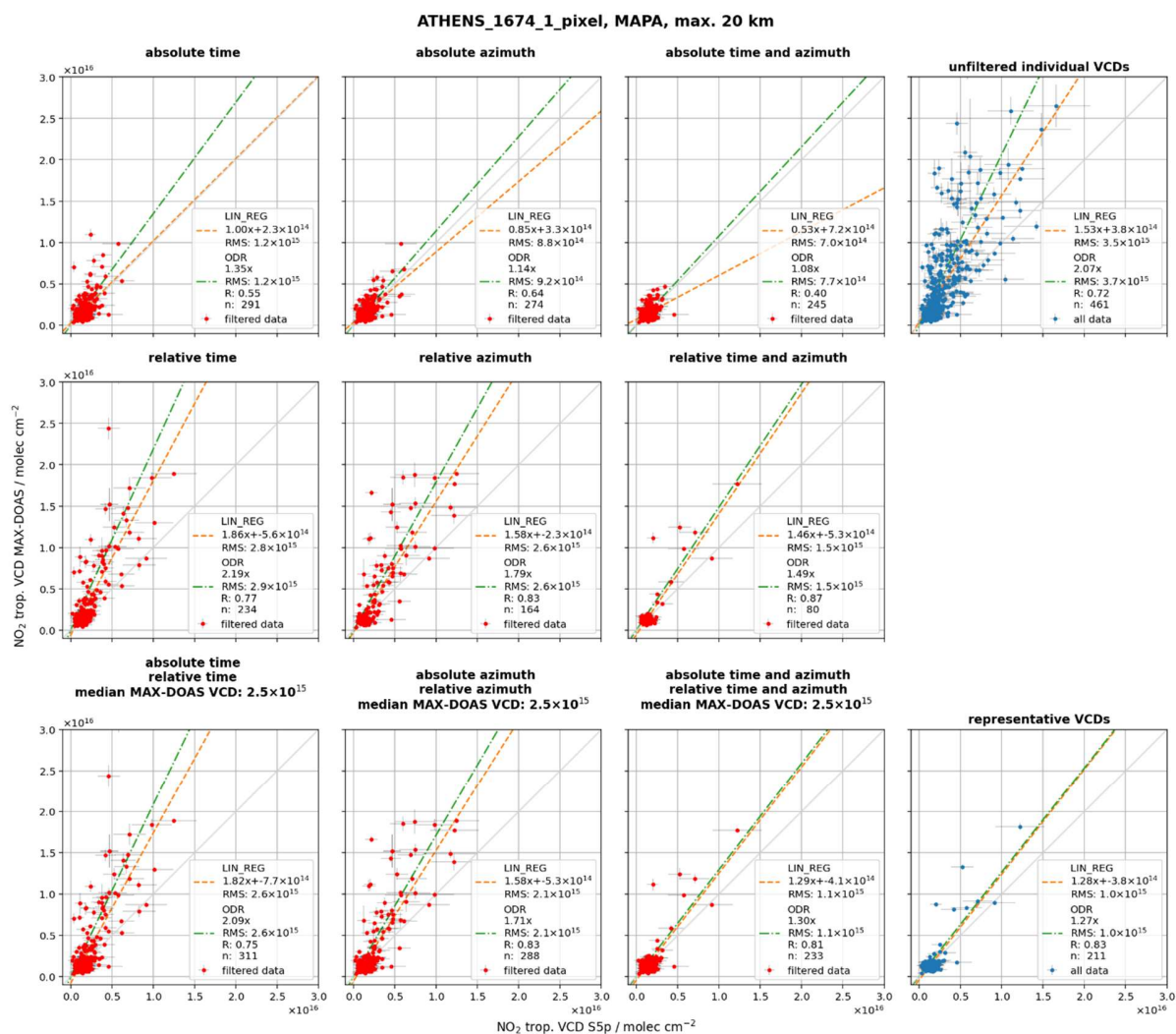


Figure 6: Scatterplots of S5p and MAPA VCDs for FRM4DOAS station Athens\_1674\_1, visible channel, with maximum distance between station and S5p pixel of 20 km.

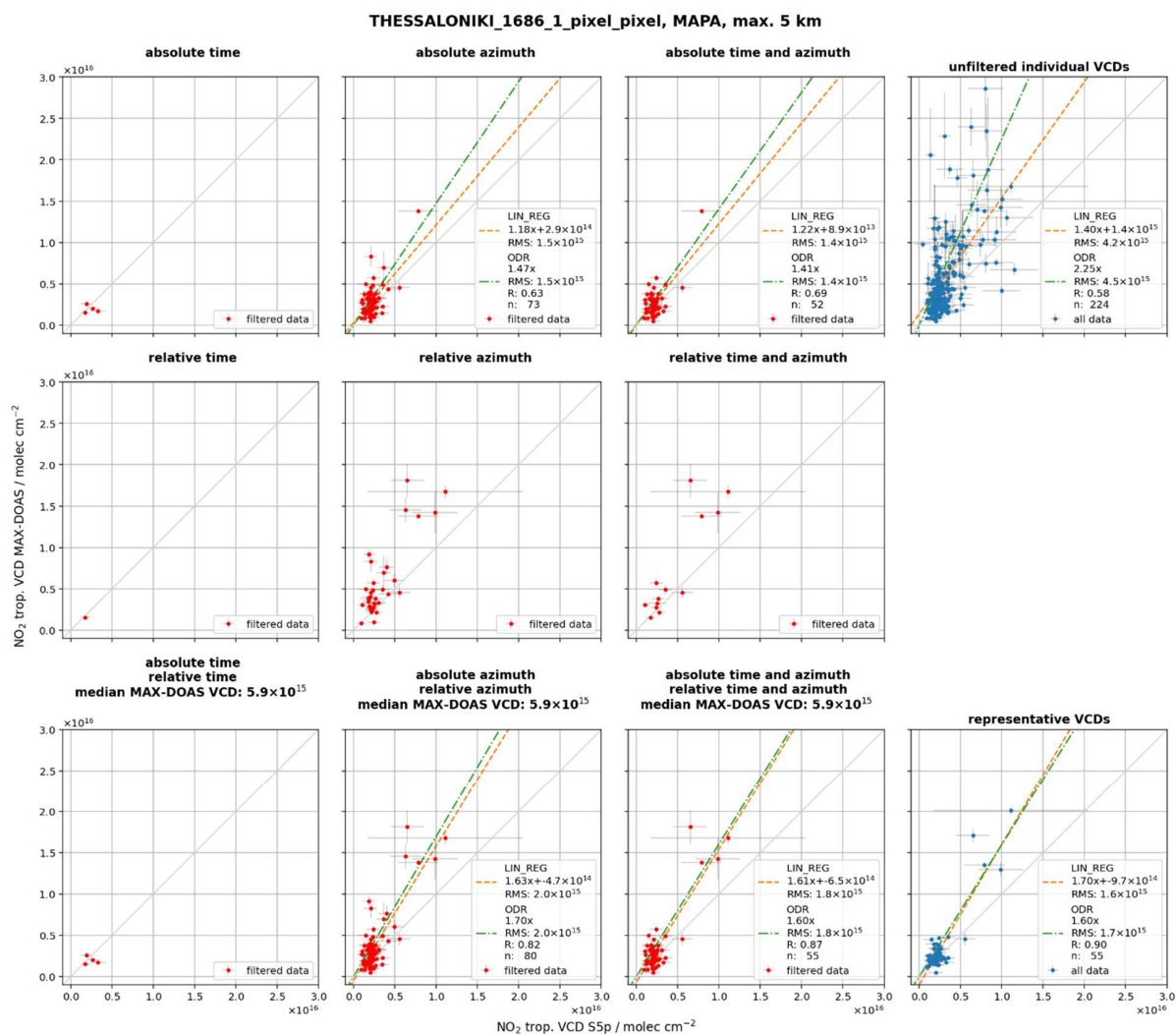


Figure 7: Scatterplots of S5p and MAPA VCDs for FRM4DOAS station Thessaloniki\_1686\_1, with a maximum distance between station and assigned S5p pixel of 20 km.

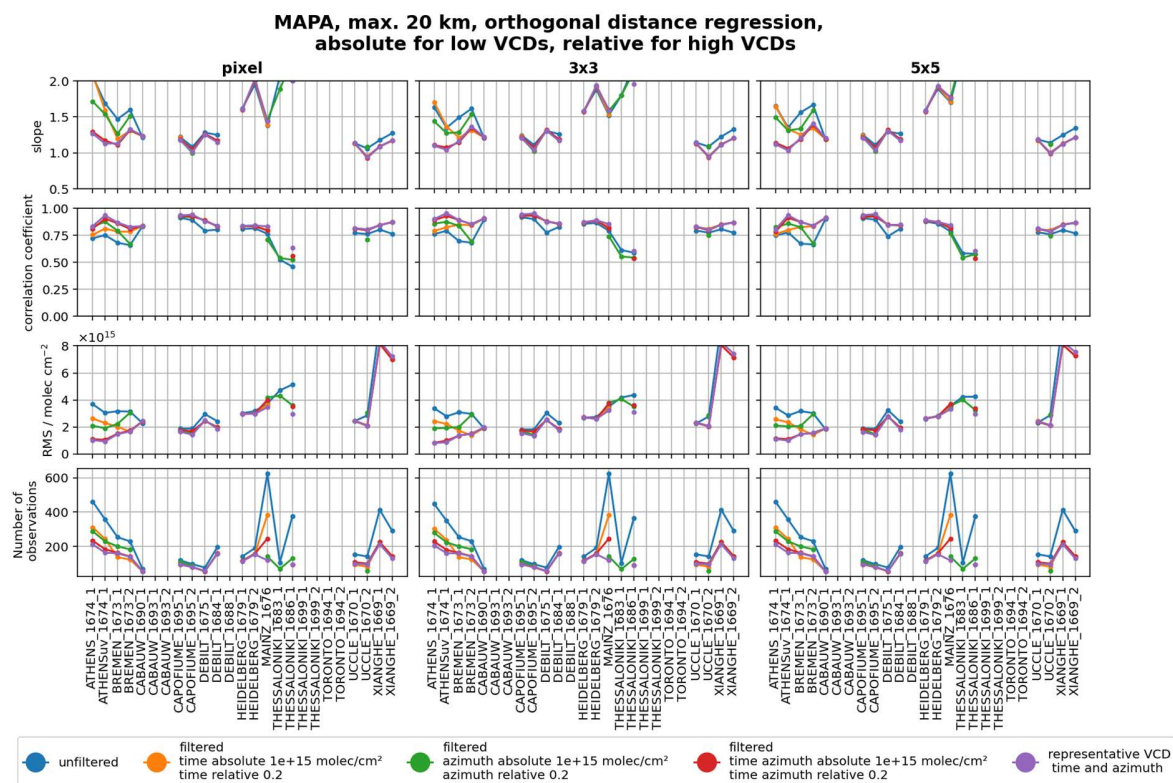


Figure 8: Overview of the results of orthogonal distance regression between MAPA VCDs and S5p pixels for all stations, with a maximum distance of 20 km. The first column shows results for individual pixels, the second column for 3x3 average of S5p pixels, and the third column for 5x5 averages.

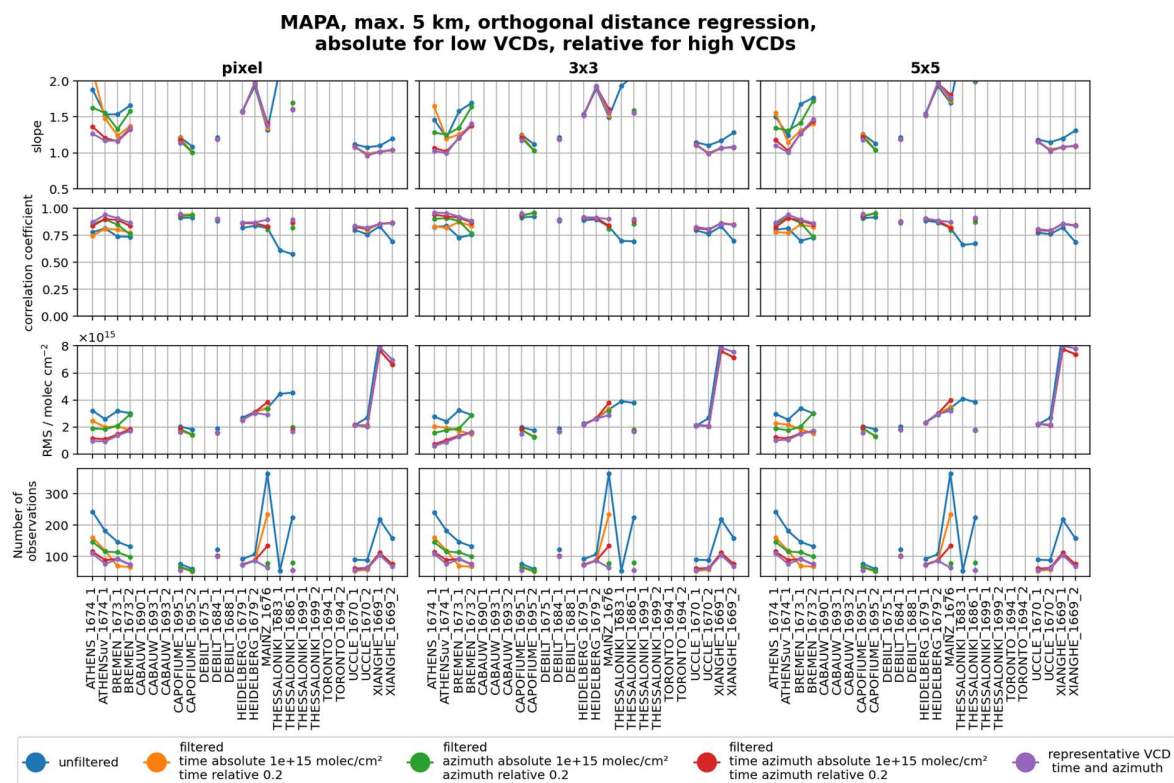


Figure 9: Overview of the results of orthogonal distance regression between MAPA VCDs and S5p pixels for all stations, with a maximum distance of 5 km. The first column shows results for individual pixels, the second column for 3x3 average of S5p pixels, and the third column for 5x5 averages.

Figure 10 displays the performance of different VCD methods across three VCDs (pixel, 3x3, 5x5) and four metrics (slope, correlation coefficient, RMS, and Number of observations). The methods are compared for absolute and azimuthal time relative to the VCD. The legend indicates: unfiltered (blue), filtered absolute time (orange), filtered azimuthal time (green), filtered time azimuthal (red), and representative VCD (purple).

The metrics are:

- slope**: Y-axis range 0.5 to 2.0.
- correlation coefficient**: Y-axis range 0.00 to 1.00.
- RMS / molec cm<sup>-2</sup>**: Y-axis range 0 to 8 (scaled by  $\times 10^{15}$ ).
- Number of observations**: Y-axis range 0 to 600.

The X-axis for all plots lists the following methods: ATHENS\_1674\_1, ATHENS\_1674\_2, BREMEN\_1673\_1, BREMEN\_1673\_2, CABAUW\_1690\_1, CABAUW\_1690\_2, CABAUW\_1693\_1, CABAUW\_1693\_2, CAPOFUME\_1695\_1, CAPOFUME\_1695\_2, DEBILT\_1684\_1, DEBILT\_1684\_2, DEBILT\_1688\_1, DEBILT\_1688\_2, HEIDELBERG\_1679\_1, HEIDELBERG\_1679\_2, MAINZ\_1676\_1, MAINZ\_1676\_2, THESSALONIKI\_1683\_1, THESSALONIKI\_1683\_2, THESSALONIKI\_1686\_1, THESSALONIKI\_1686\_2, THESSALONIKI\_1699\_1, THESSALONIKI\_1699\_2, TORONTO\_1694\_1, TORONTO\_1694\_2, UCCLLE\_1670\_1, UCCLLE\_1670\_2, XIANGHE\_1669\_1, XIANGHE\_1669\_2.

### 3.2 MMF

The validation results for all stations using orthogonal distance regression and a maximum distance of 20 km are shown in Figure 12. Also here, filtering for the combined inhomogeneity of time and azimuth increases the correlation coefficients and decreases the RMS. In all cases, the regressions using the representative columns show better results than the individually filtered VCDs, even though for some stations the difference is rather small. Averaging of S5p pixels leads to station specific results, for Athens and Bremen the slopes are closer to 1, while



for Mainz the slope increases and for Cabauw, San Pietro Capofiume, Toronto, Uccle and Xianghe there is almost no difference.

A reduction of distance between station and S5p pixel center to 5 km further increases the correlation coefficients and leads to a small reduction in the RMS, and also here the representative columns show the best results.

Using linear regressions (Figure 14), the correlation coefficients also increase, and the RMS is reduced as well, while the influence on the slope depends on the station. For some stations, filtering improves the slope, while for others, the unfiltered data show better results.

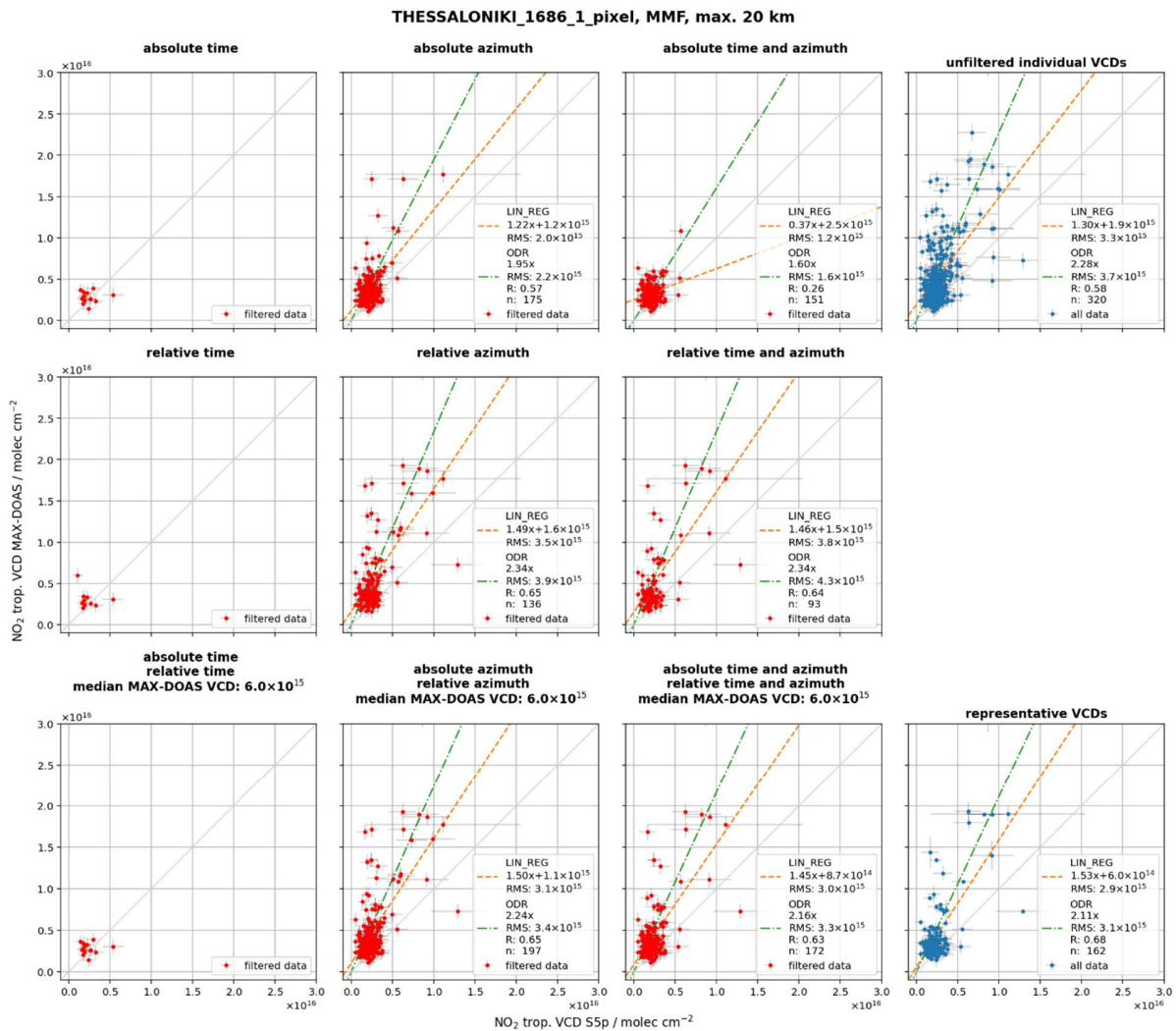


Figure 11: Scatterplots of S5p and MMF VCDs for FRM4DOAS station Thessaloniki\_1686\_1, with a maximum distance between station and assigned S5p pixel of 20 km.

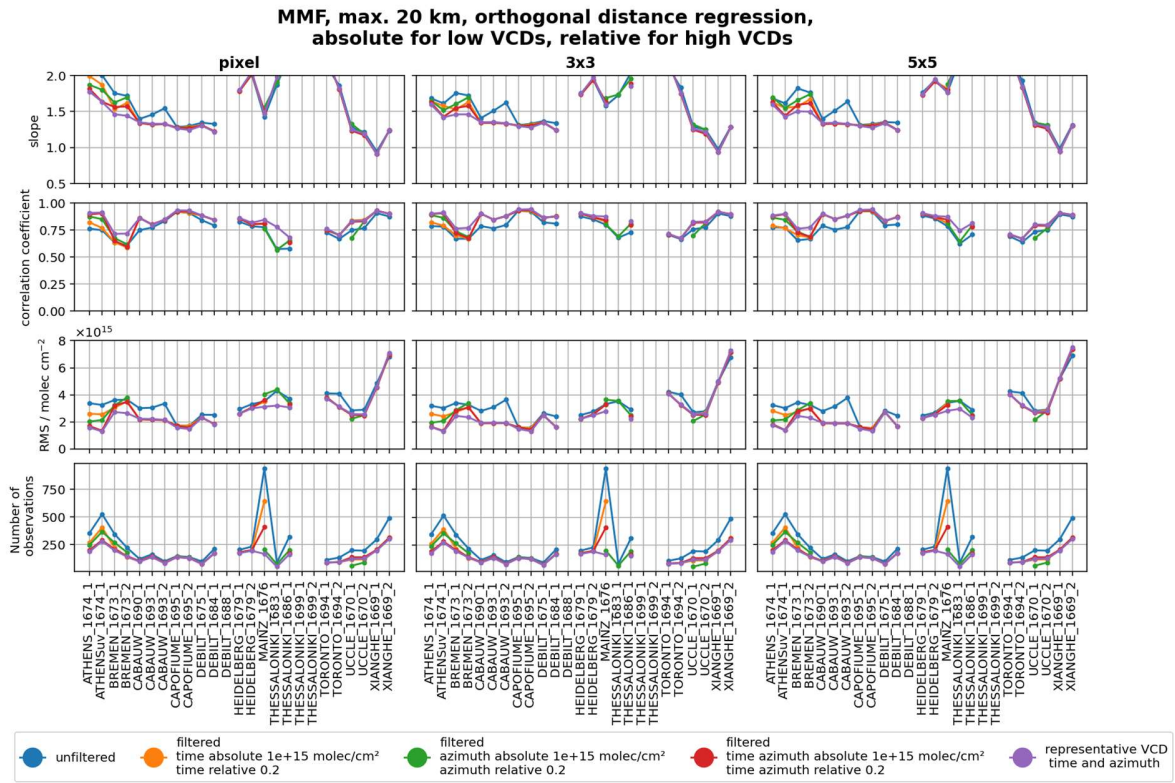


Figure 12: Overview of the results of orthogonal distance regression between MMF VCDs and S5p pixels for all stations, with a maximum distance of 20 km. The first column shows results for individual pixels, the second column for 3x3 averages of S5p pixels, and the third column for 5x5 averages.

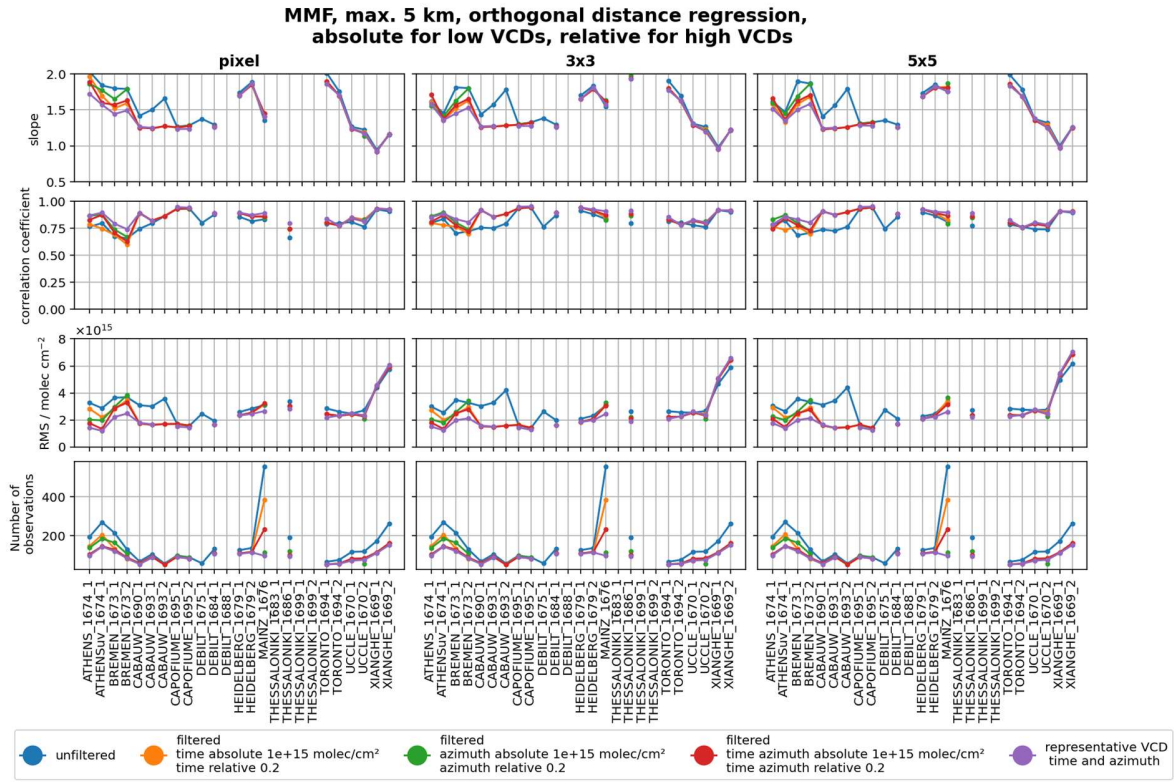


Figure 13: Overview of the results of orthogonal distance regression between MMF VCDs and S5p pixels for all stations, with a maximum distance of 5 km. The first column shows results for individual pixels, the second column for 3x3 averages of S5p pixels, and the third column for 5x5 averages.



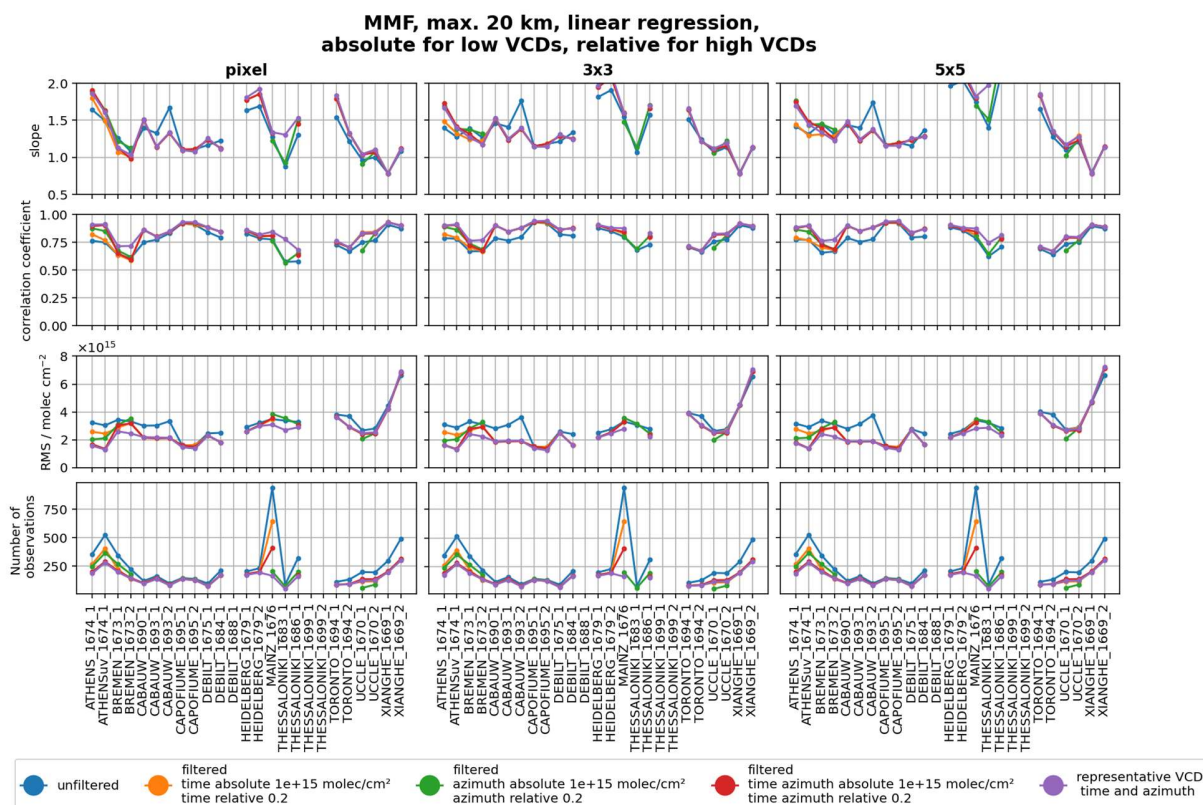


Figure 14: Overview of the results of linear regression between MMF VCDs and S5p pixels for all stations, with a maximum distance of 20 km. The first column shows results for individual pixels, the second column for 3x3 averages of S5p pixels, and the third column for 5x5 averages.

## 4 Summary

Three different NO<sub>2</sub> inhomogeneity indicators were defined and calculated for several urban MAX-DOAS stations. The inhomogeneity indicators indicate azimuthal, temporal, and the combined azimuthal and temporal inhomogeneity in the NO<sub>2</sub> field around a station, based on the tropospheric VCDs derived by the MAPA and MMF algorithms. All inhomogeneity indicators were defined as absolute inhomogeneity (in molec/cm<sup>2</sup>) and as relative inhomogeneity, where a median was used for normalization.

MAPA and MMF tropospheric NO<sub>2</sub> VCDs have been matched to the respective S5p tropospheric NO<sub>2</sub> VCDs, and their agreement was investigated using orthogonal distance regression and linear regression between the MAX-DOAS and corresponding S5p data. Matching was carried out using two different distance settings (5 km and 20 km maximum distance between station and pixel) and three different pixel settings (individual pixels, 3x3 averages, and 5x5 averages around the corresponding S5p pixel). Regressions were performed on the unfiltered data pairs, and on filtered data pairs, where the data has been filtered either using an absolute, a relative,

or a combined inhomogeneity criterion. Additionally, regressions were performed using a representative column, which is the mean of several filtered VCDs with different azimuth directions.

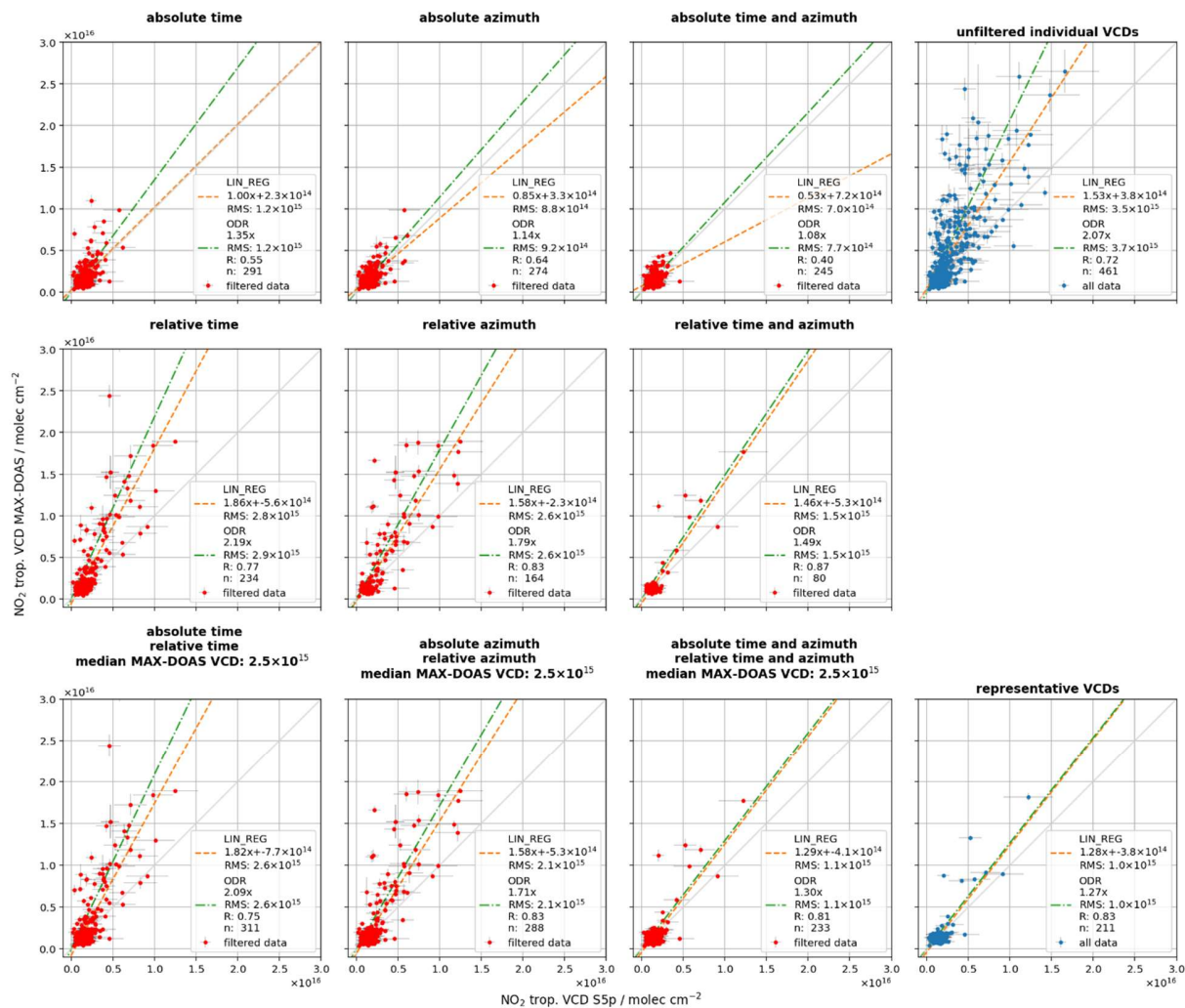
For most cases, filtering for inhomogeneity improves the correlation coefficients and decreases the RMS between MAX-DOAS VCDs and corresponding S5p VCDs. Filtering for the combined temporal and azimuthal inhomogeneity using a combination of absolute criterion for VCDs smaller than the station's median VCD and a relative criterion for VCDs bigger than the station's median VCD shows the best results. The common filter settings generally worked well for all stations, but for individual stations, custom settings could further improve the results, especially when using only absolute indicators. Correlations of individually filtered VCDs and representative averaged VCDs that have been filtered for the same thresholds using the same indicators show similar results, but in most cases the results using the representative VCDs are better.

For potential users we suggest to use the combined temporal and azimuthal inhomogeneity indicators and also to combine relative and absolute thresholds, filtering for absolute indicators when the VCDs are low and relative when the VCDs are high. The differentiation between low and high VCDs should be carried out for each station individually e.g., using a station specific median VCD calculated from a long timeseries, preferably out of all data. Averaging of satellite pixels improves the correlations in most cases, but for individual stations this leads to an increase in the slope further away from 1.

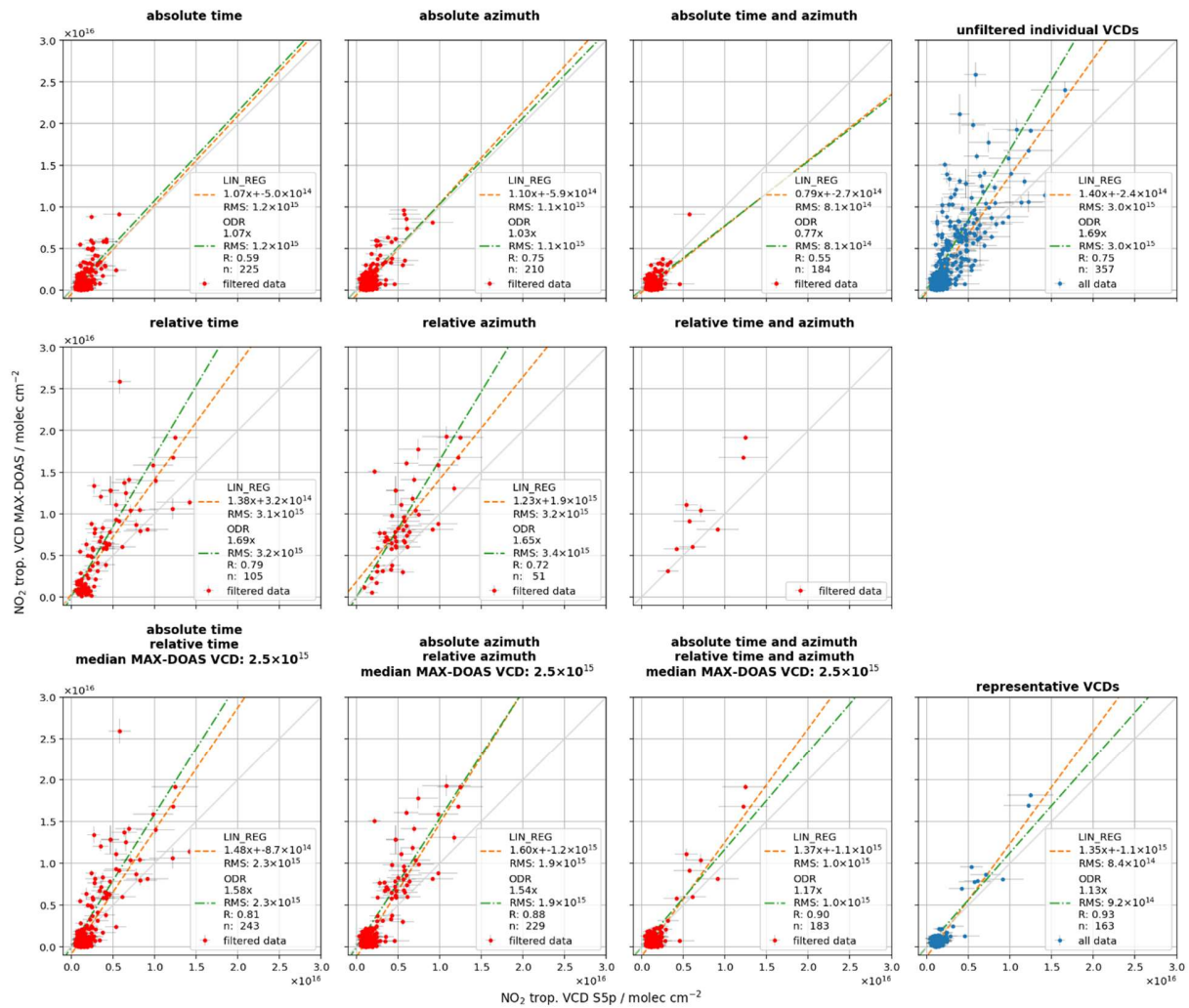
## Appendix

### Scatterplots MAPA

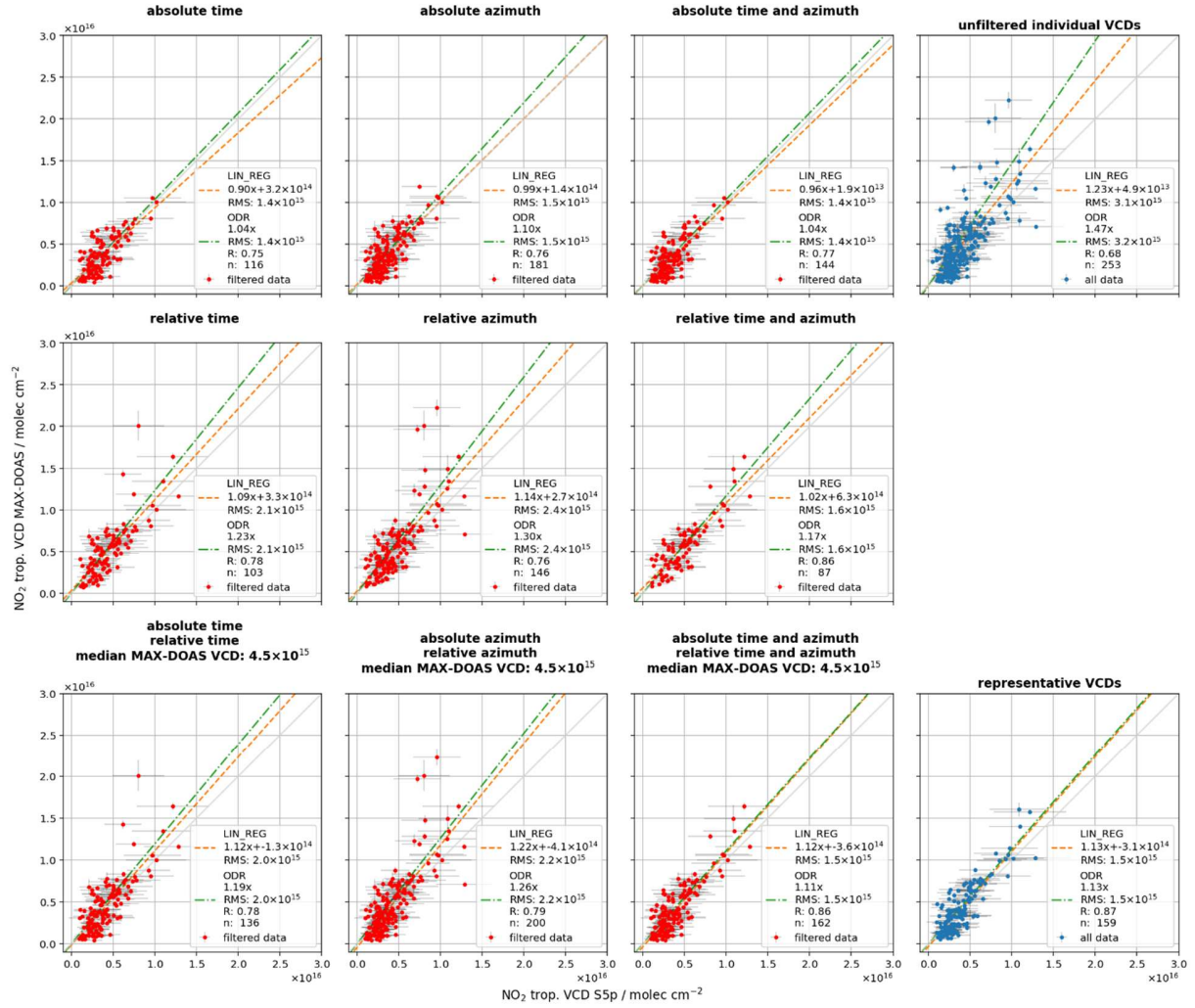
# ATHENS\_1674\_1\_pixel, MAPA, max. 20 km



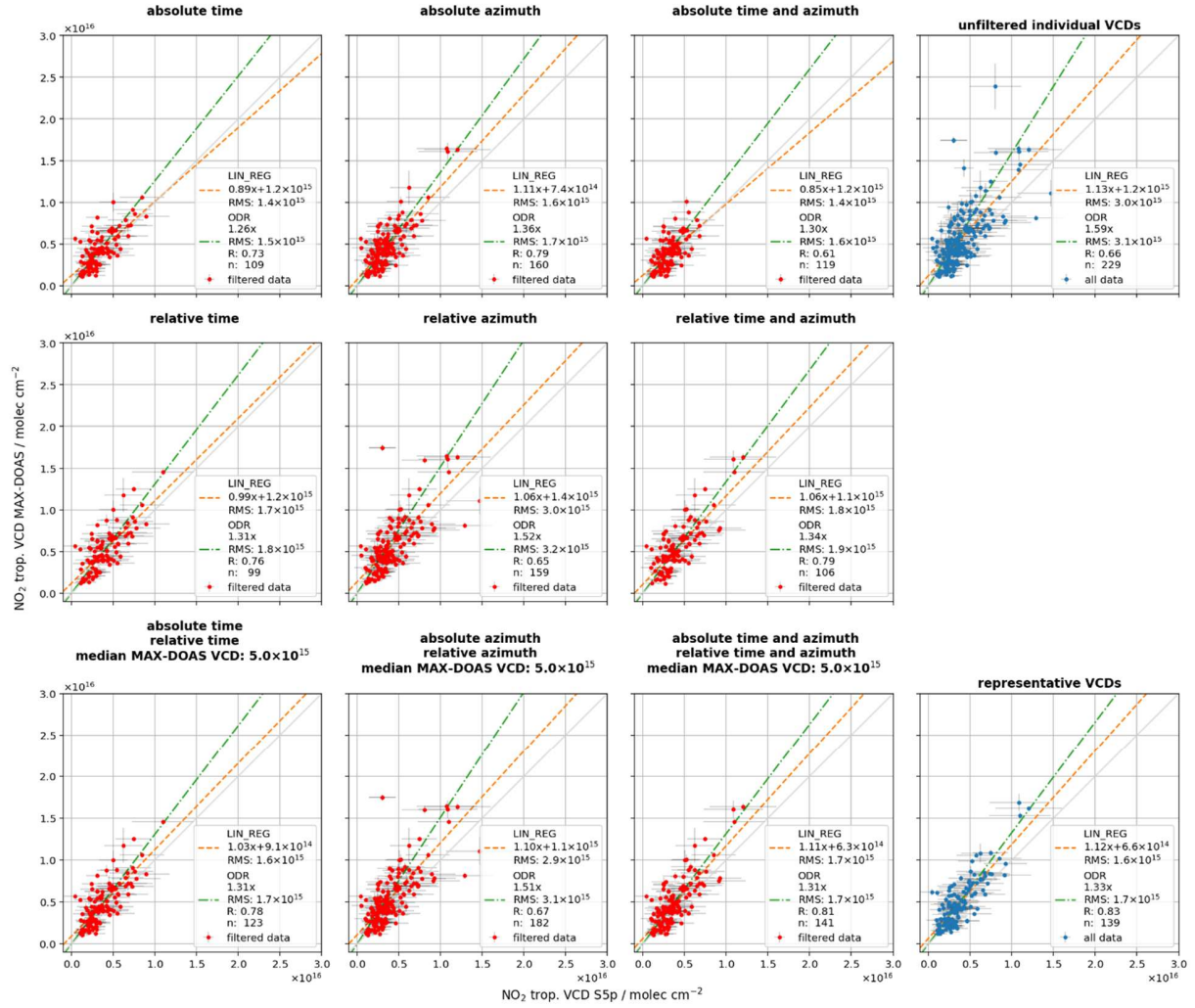
# ATHENSuv\_1674\_1\_pixel, MAPA, max. 20 km



# BREMEN\_1673\_1\_pixel, MAPA, max. 20 km

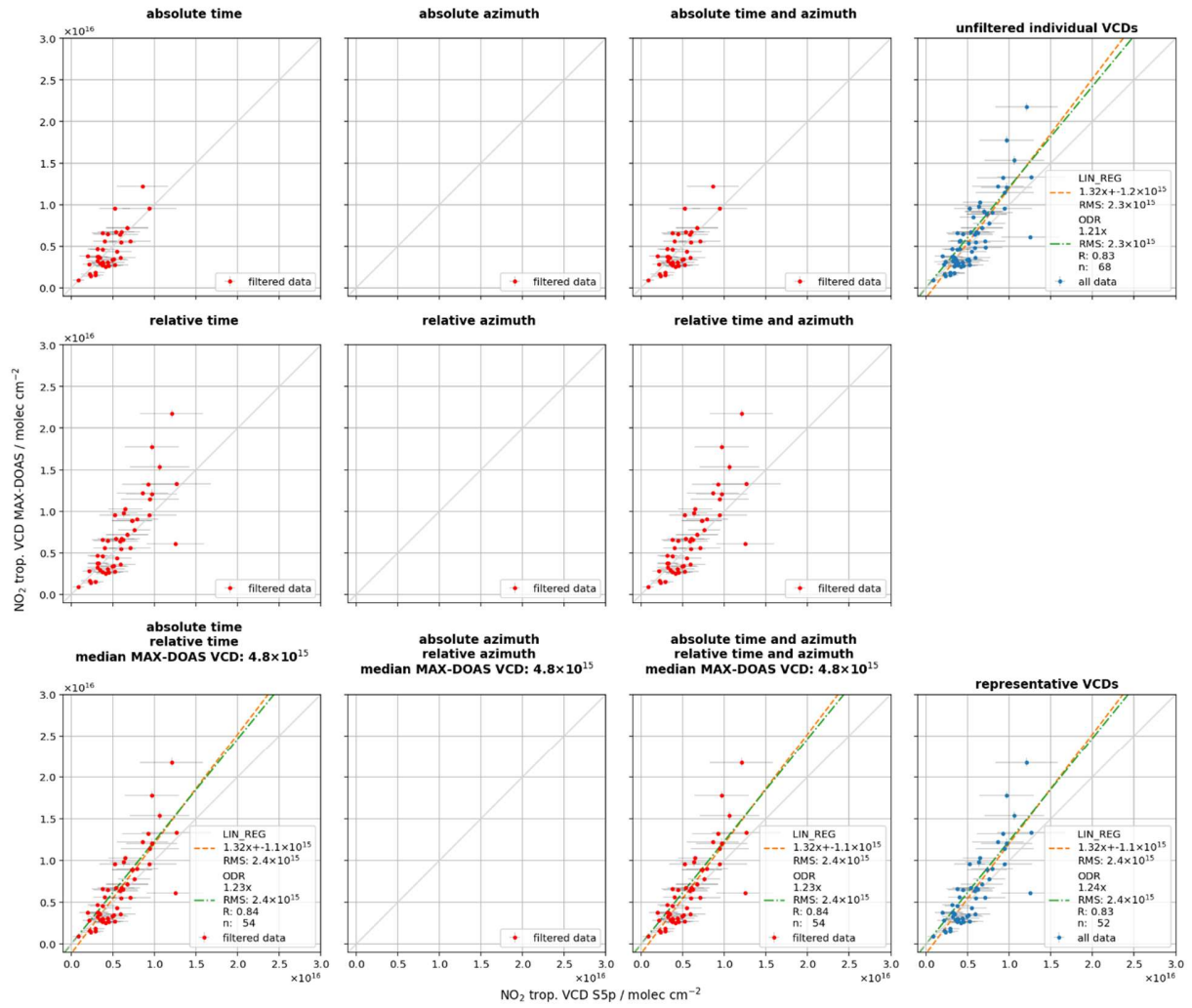


# BREMEN\_1673\_2\_pixel, MAPA, max. 20 km

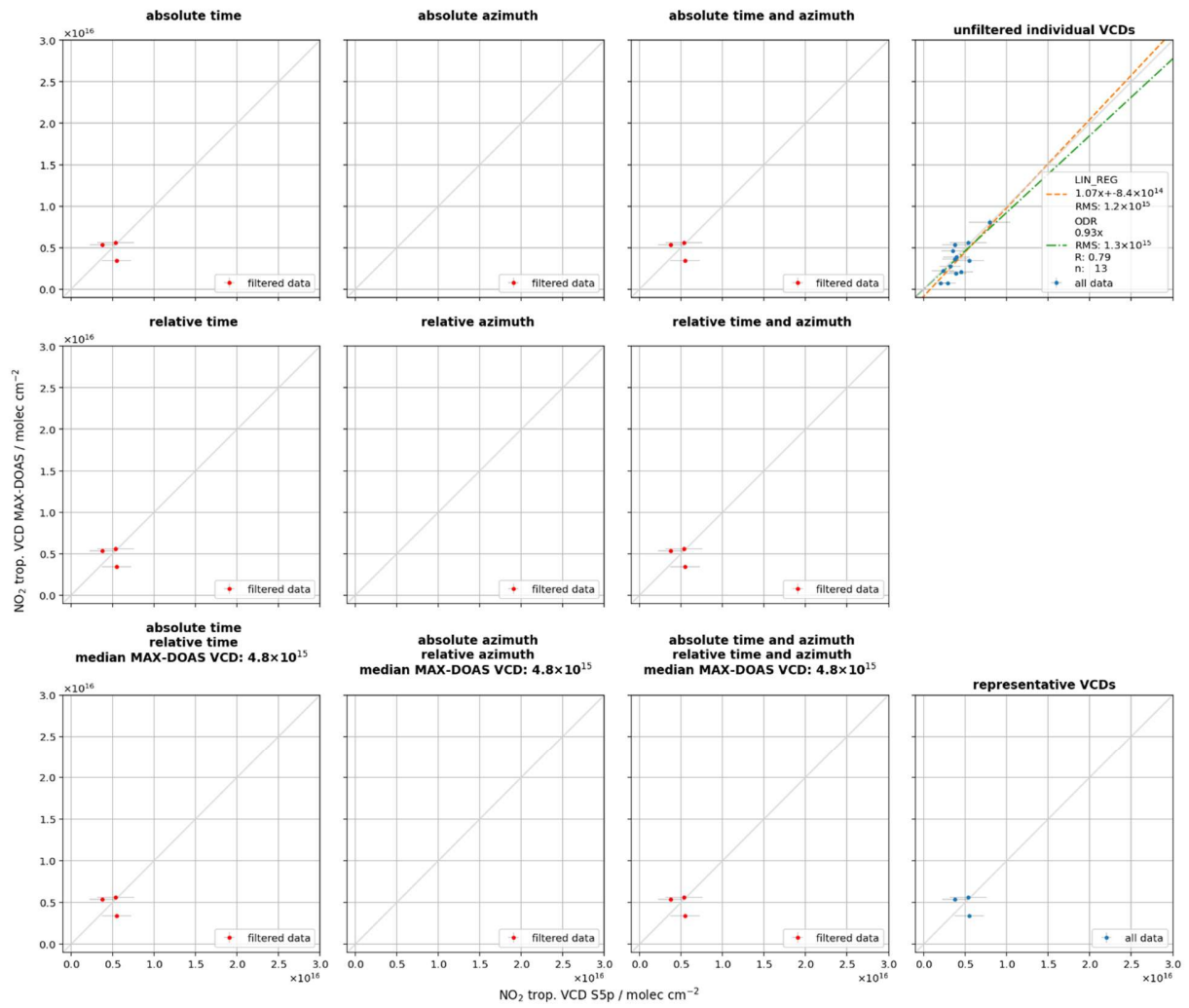




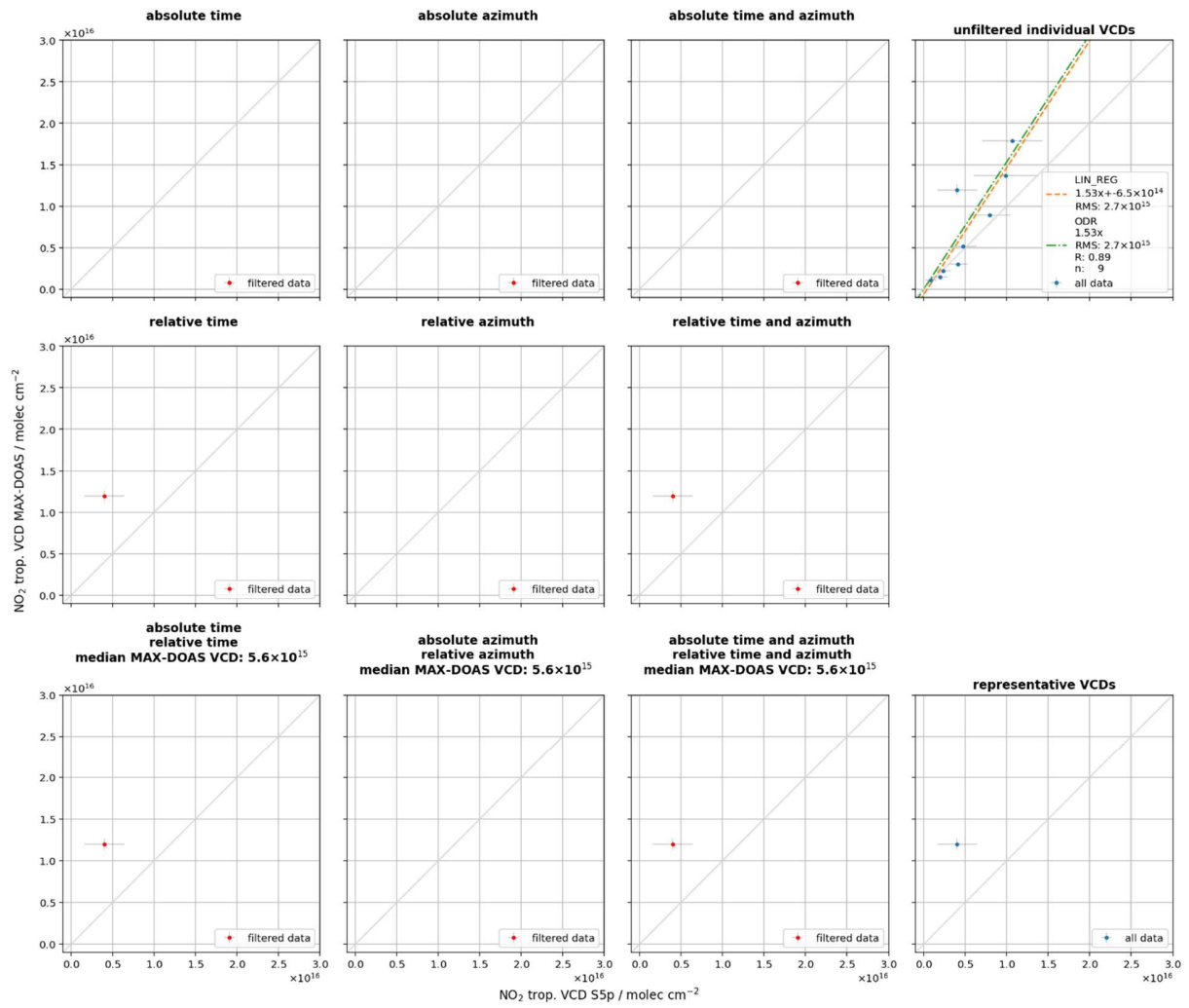
CABAUW\_1690\_1\_pixel, MAPA, max. 20 km



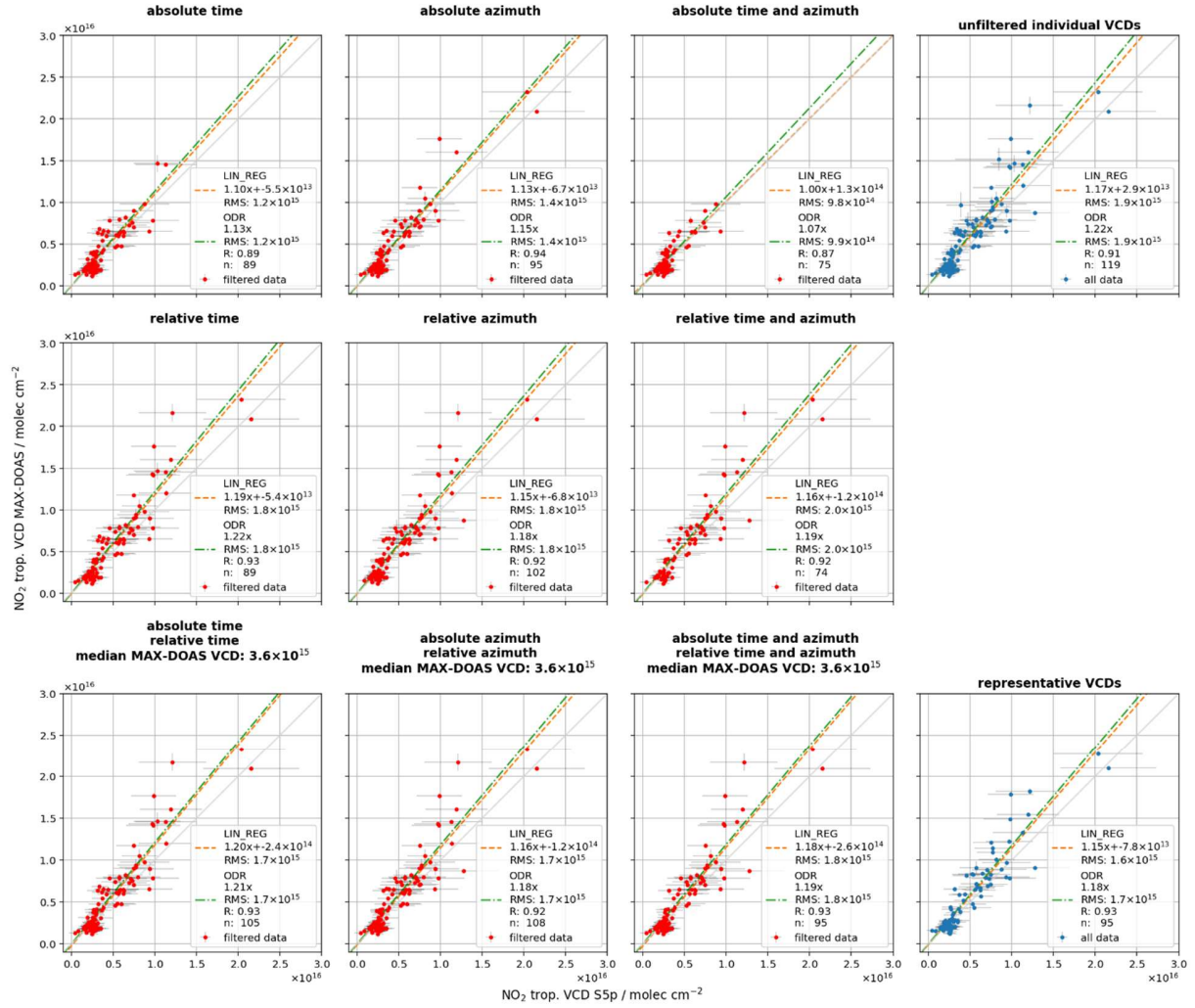
CABAUW\_1693\_1\_pixel, MAPA, max. 20 km



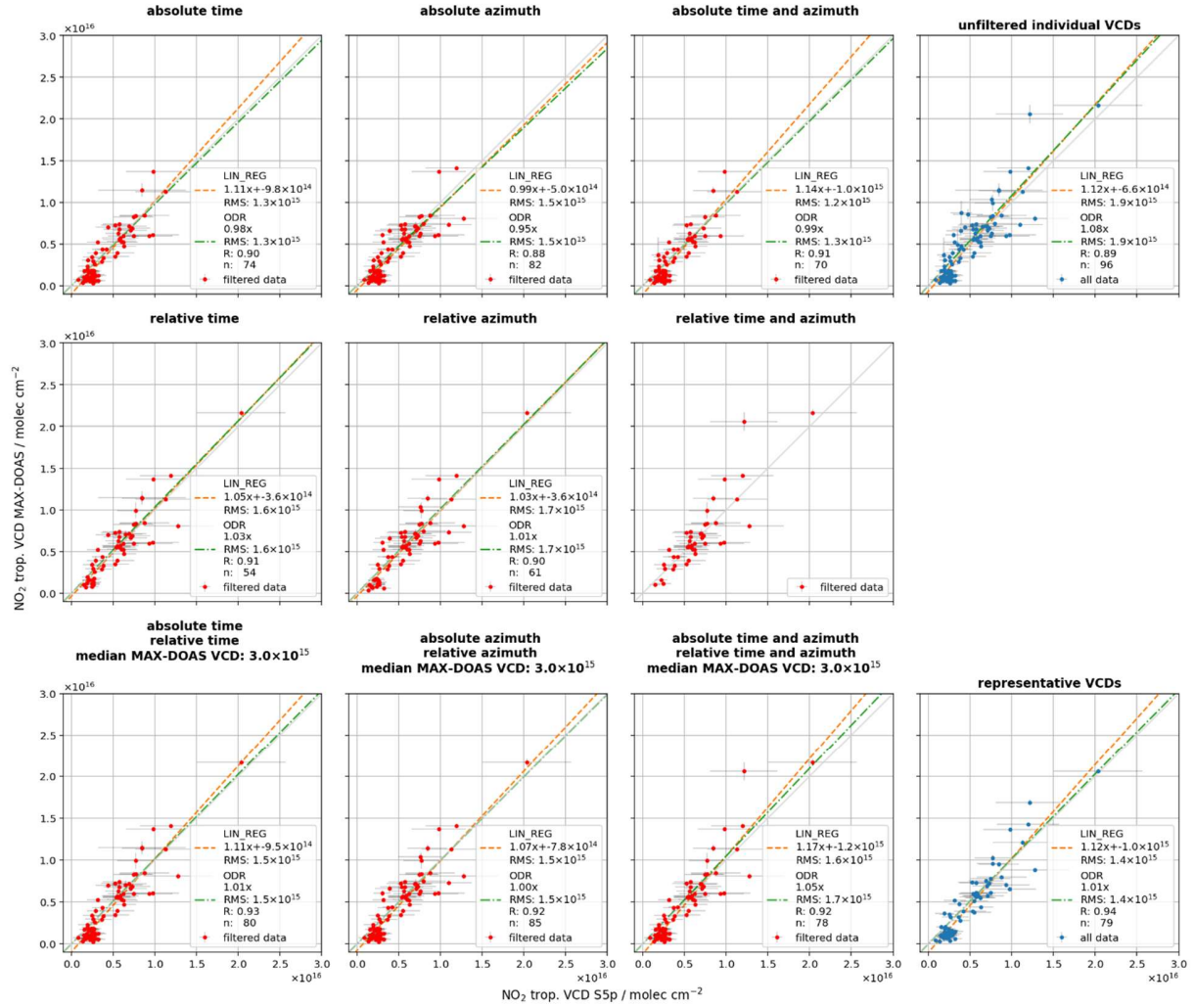
CABAUW\_1693\_2\_pixel, MAPA, max. 20 km



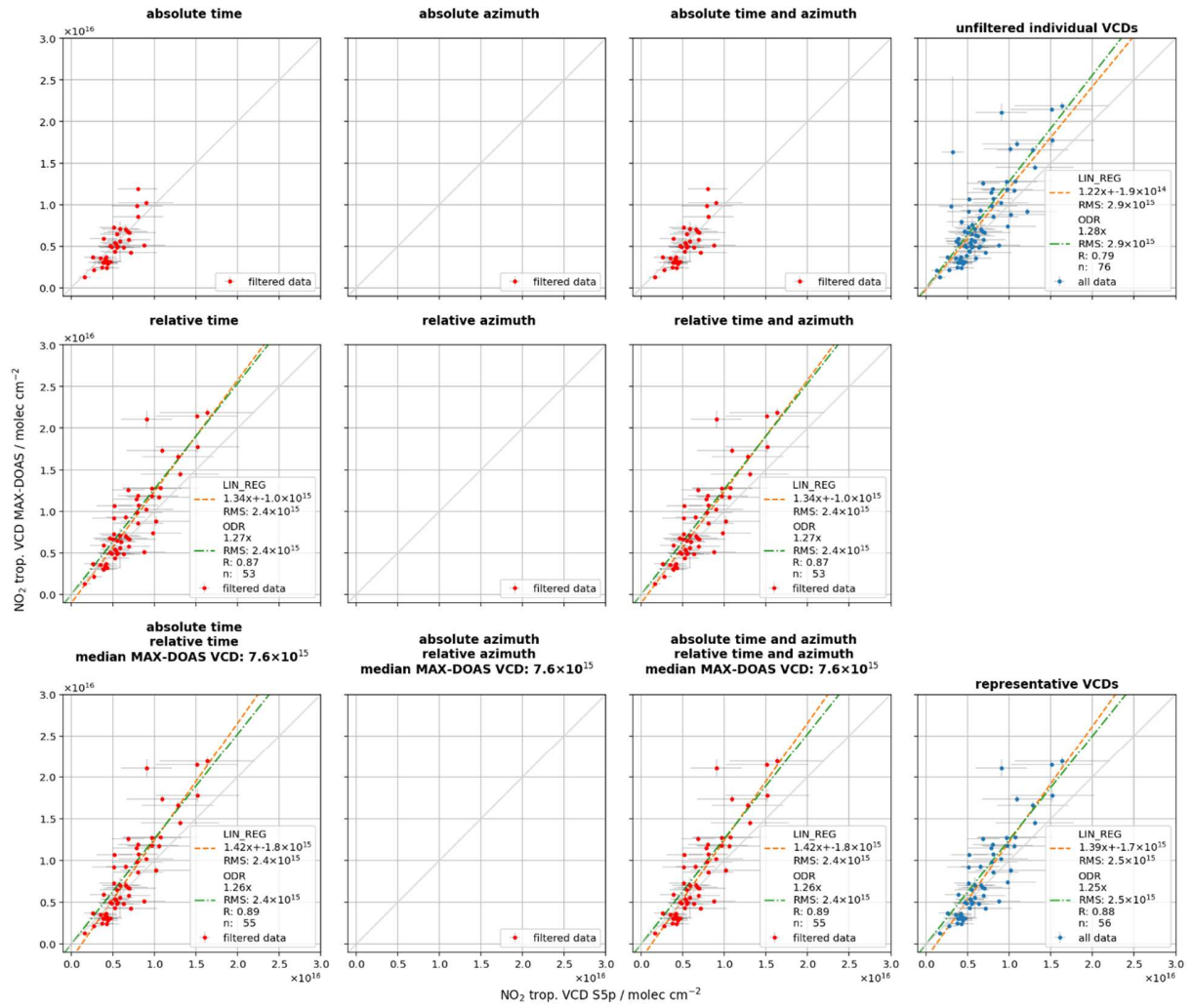
CAPOFIUME\_1695\_1\_pixel, MAPA, max. 20 km



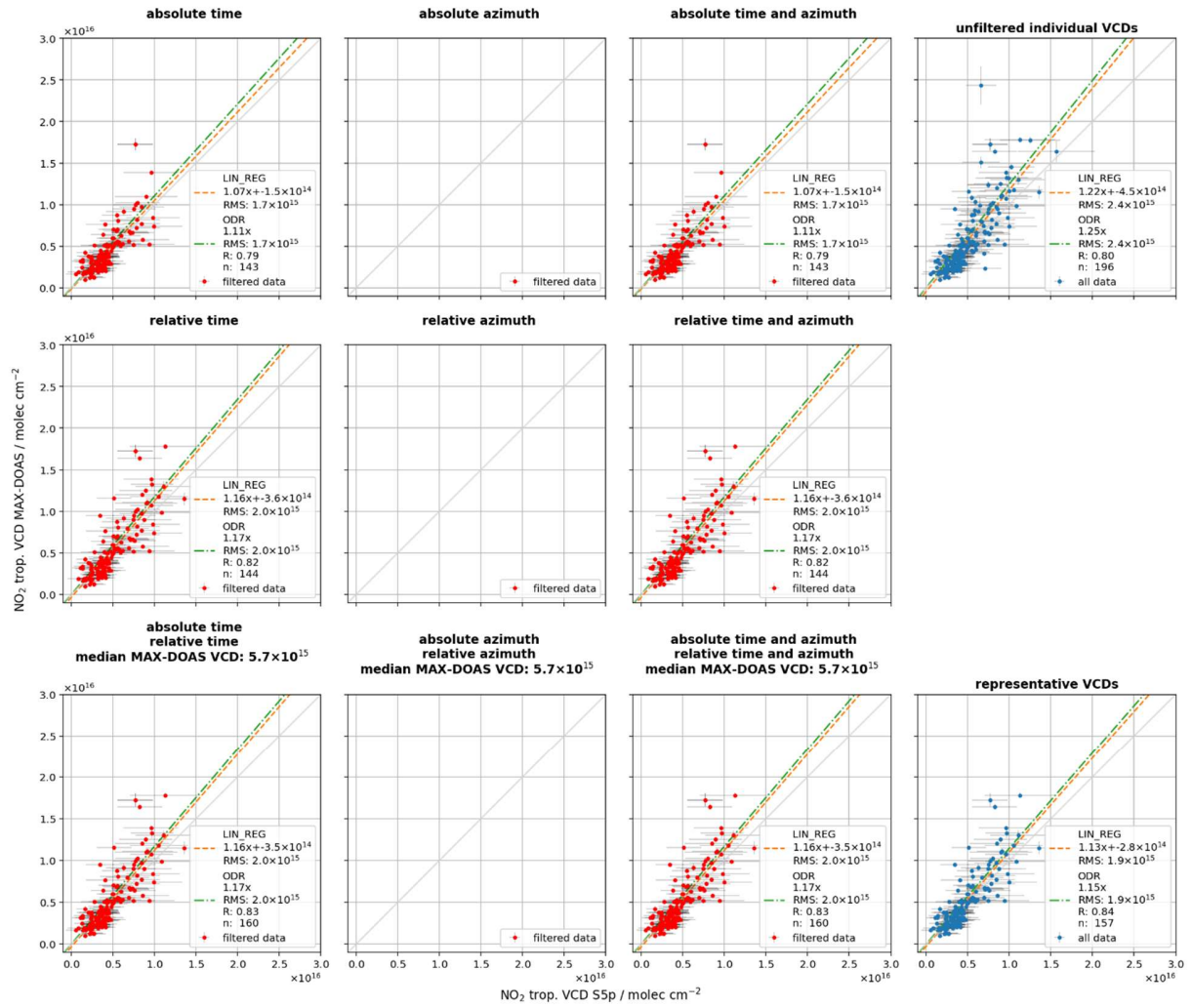
CAPOFIUME\_1695\_2\_pixel, MAPA, max. 20 km



DEBILT\_1675\_1\_pixel, MAPA, max. 20 km

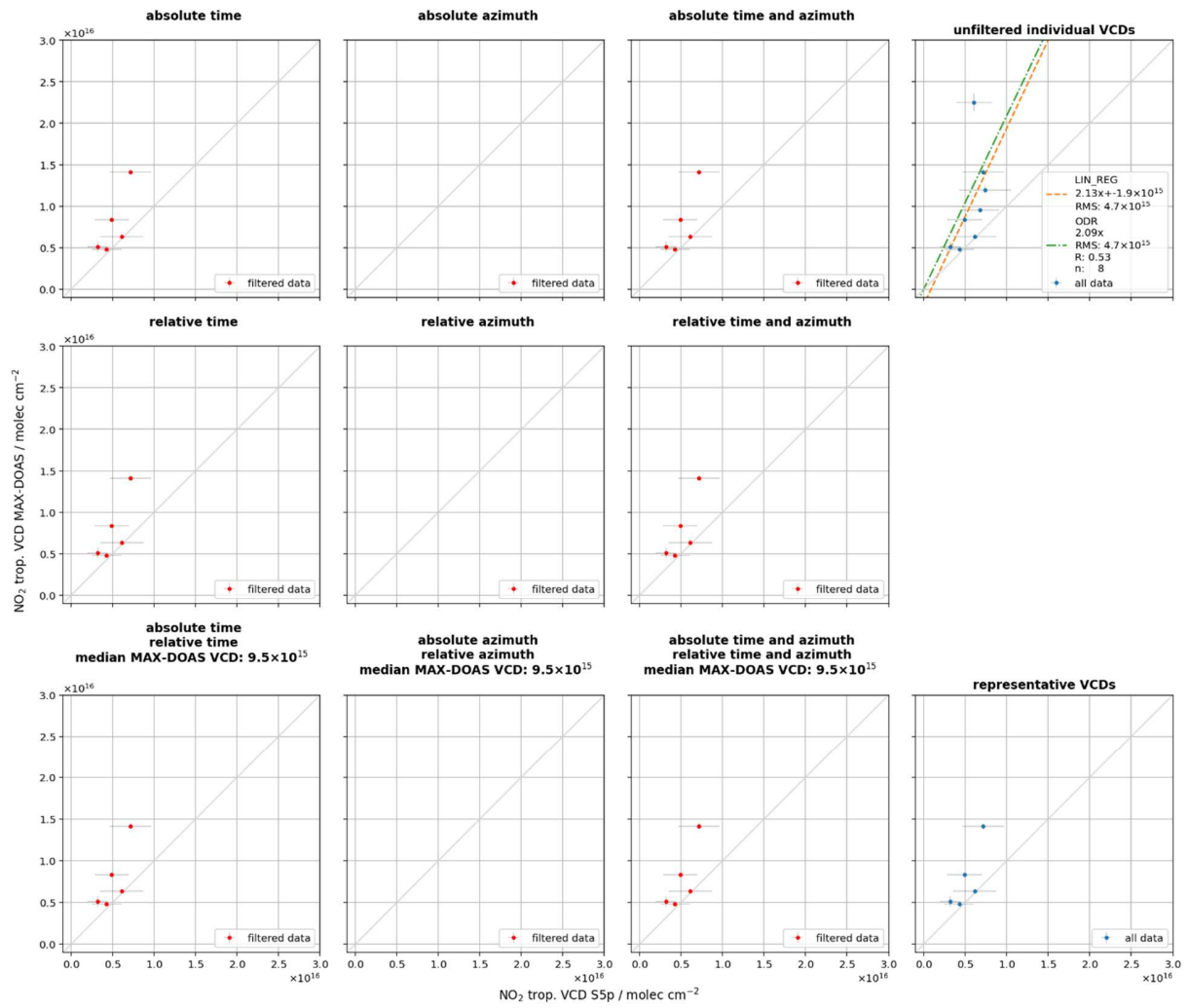


DEBILT\_1684\_1\_pixel, MAPA, max. 20 km

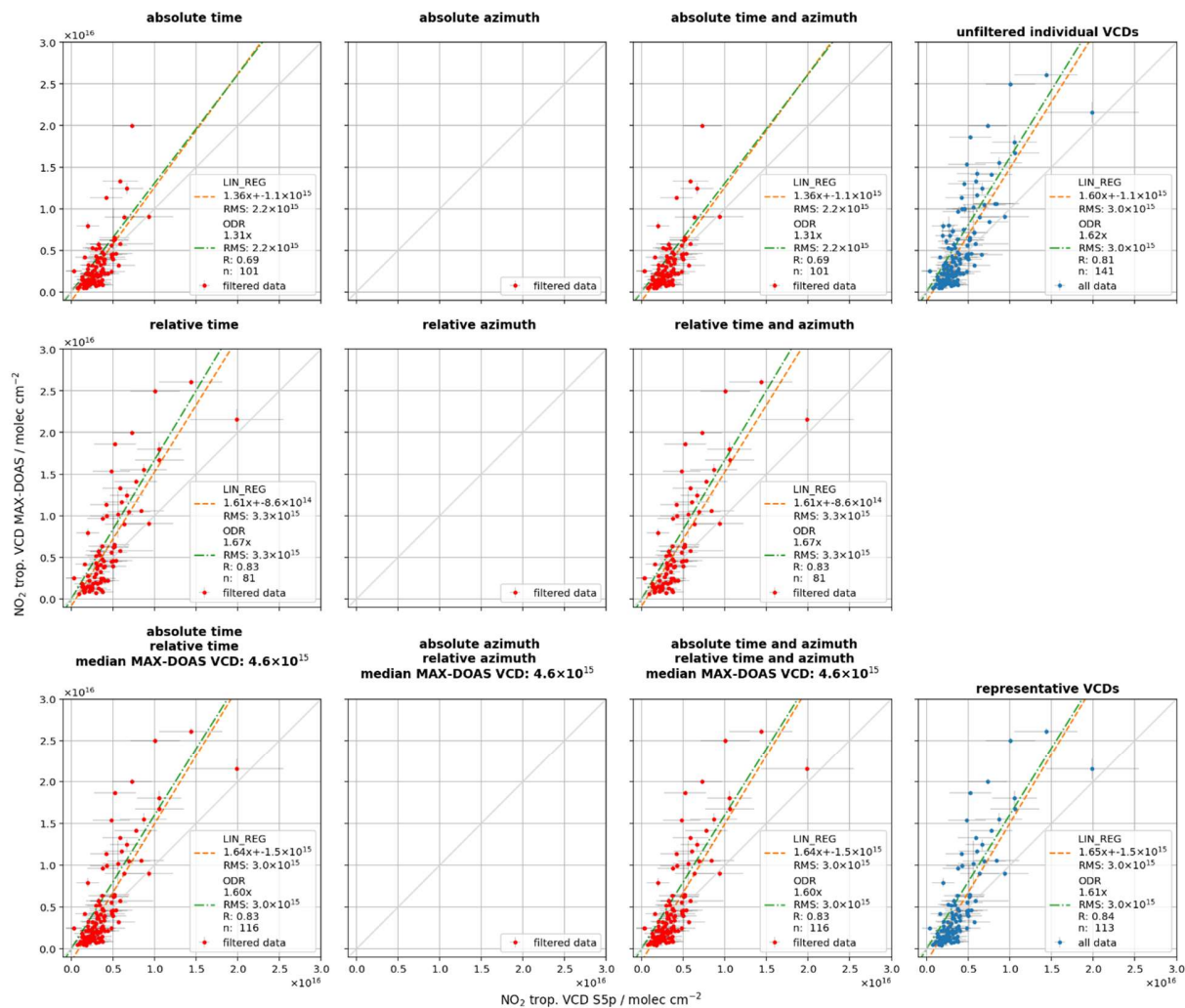




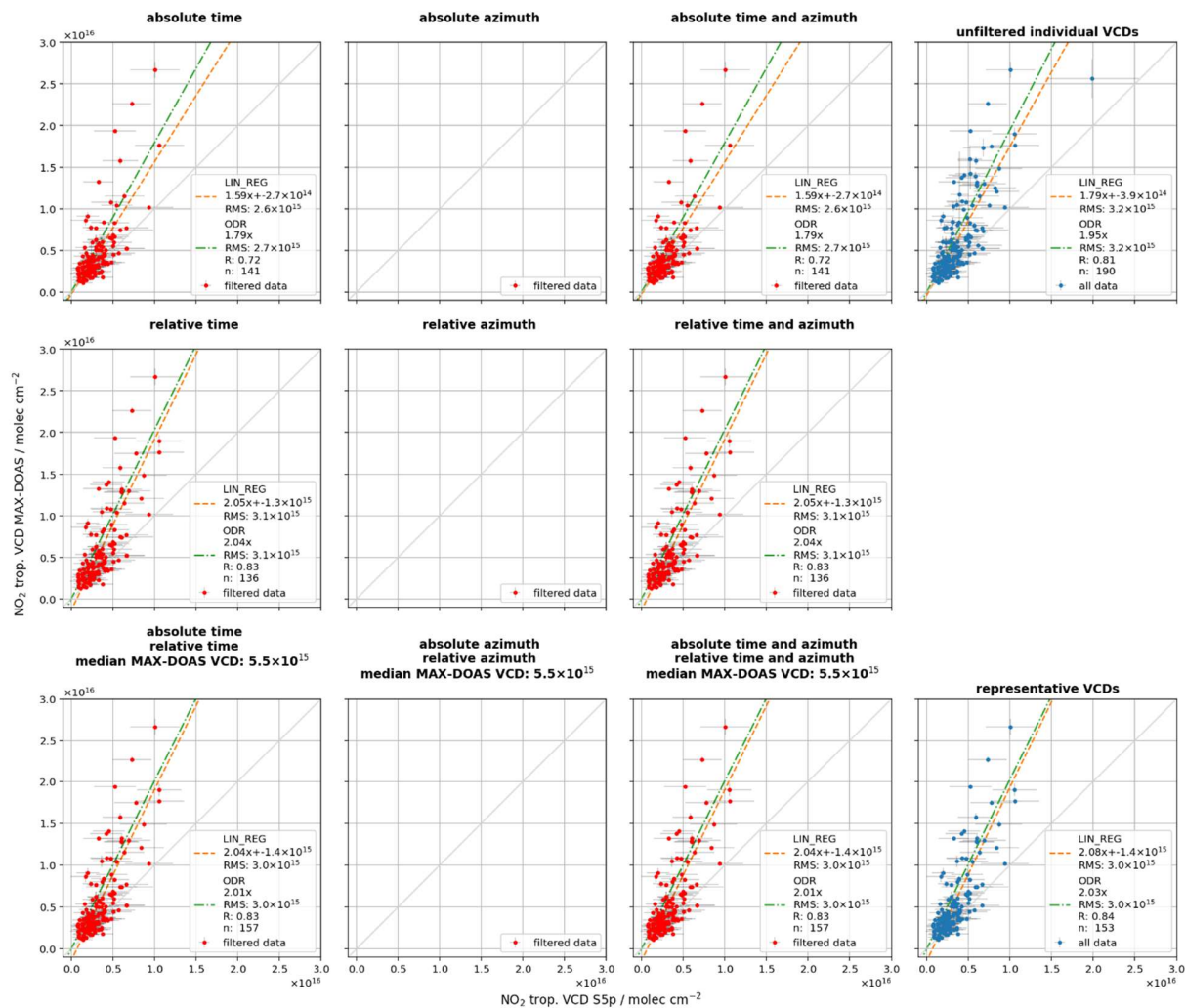
DEBILT\_1688\_1\_pixel, MAPA, max. 20 km



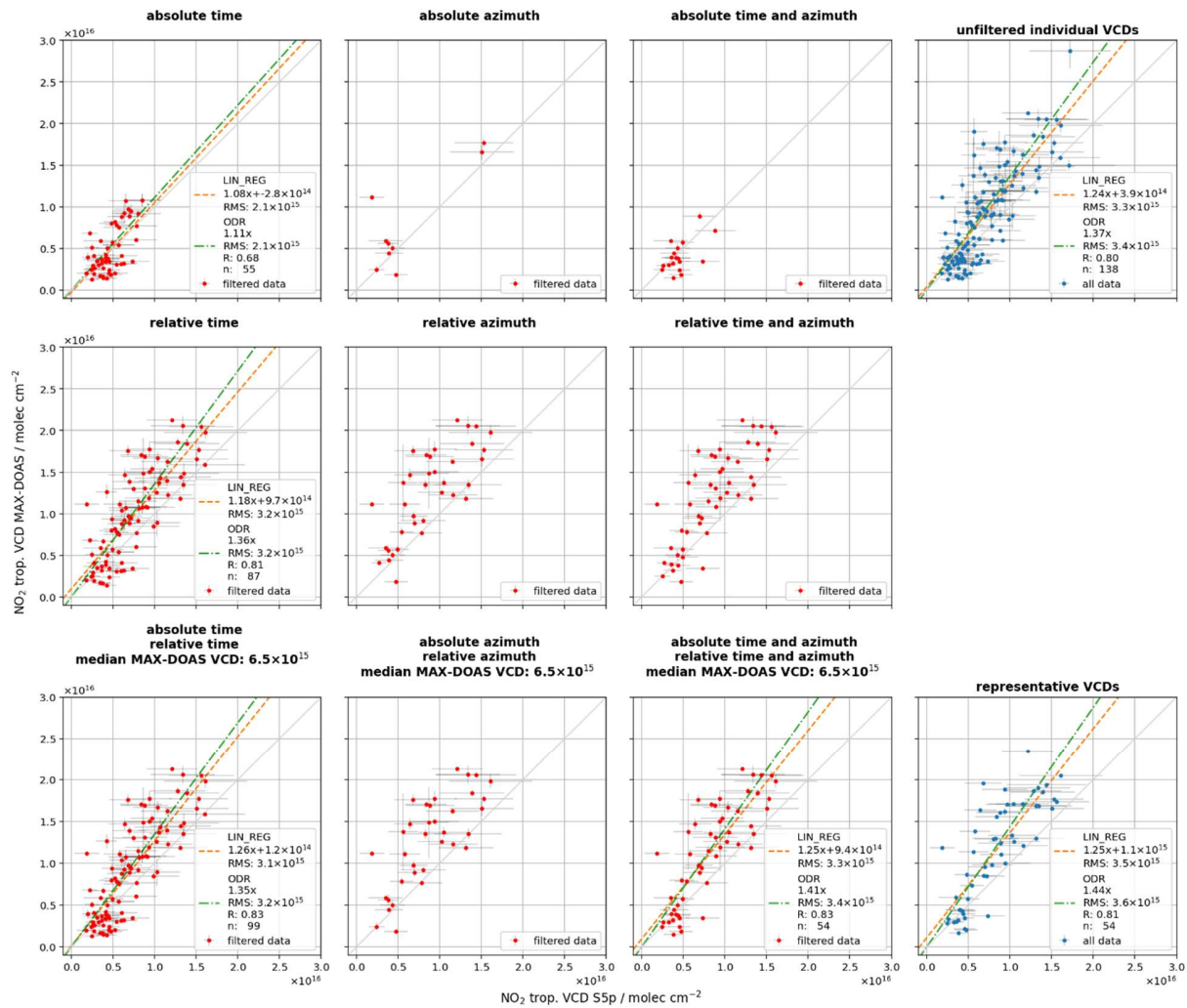
# HEIDELBERG\_1679\_1\_pixel, MAPA, max. 20 km



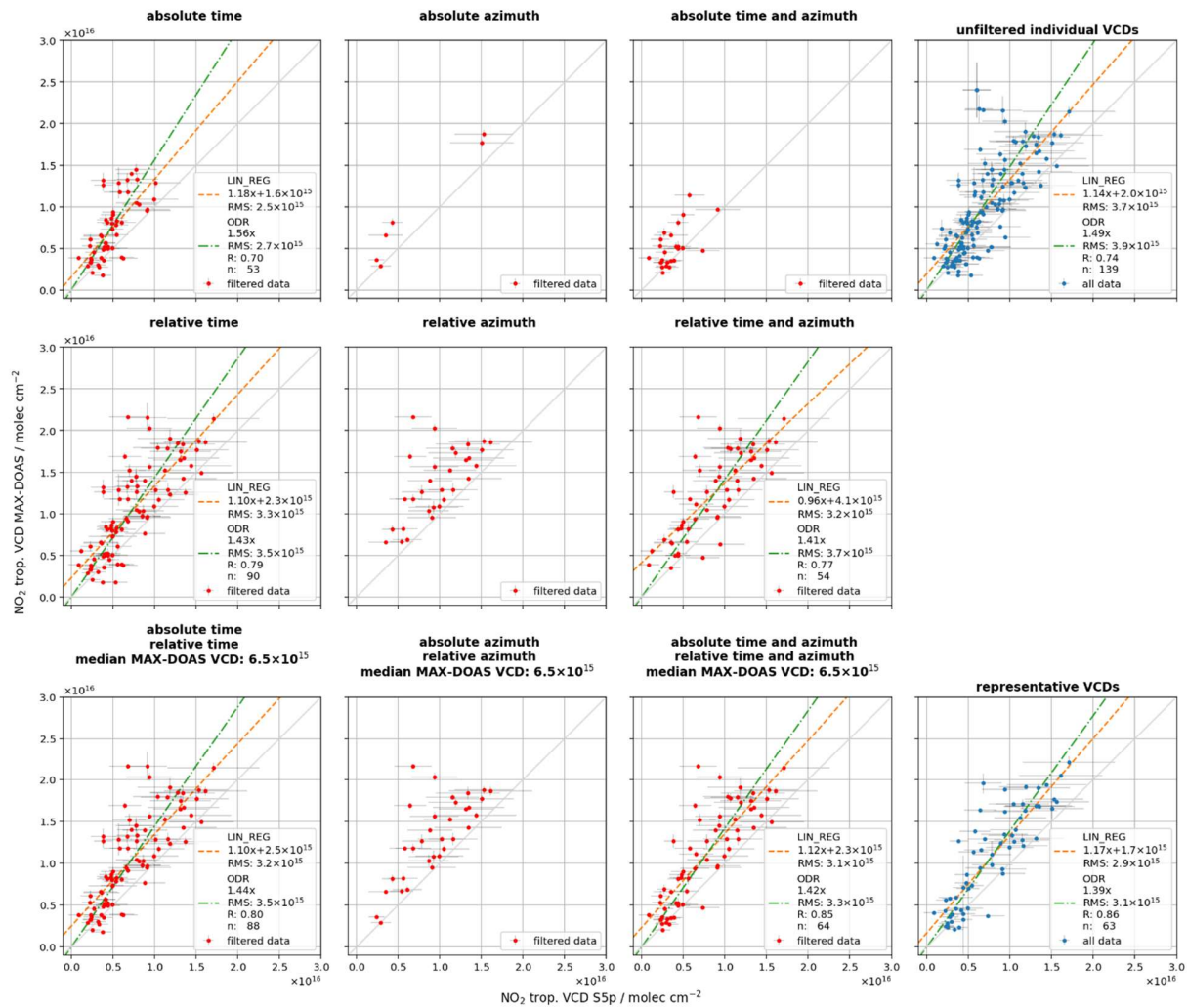
# HEIDELBERG\_1679\_2\_pixel, MAPA, max. 20 km



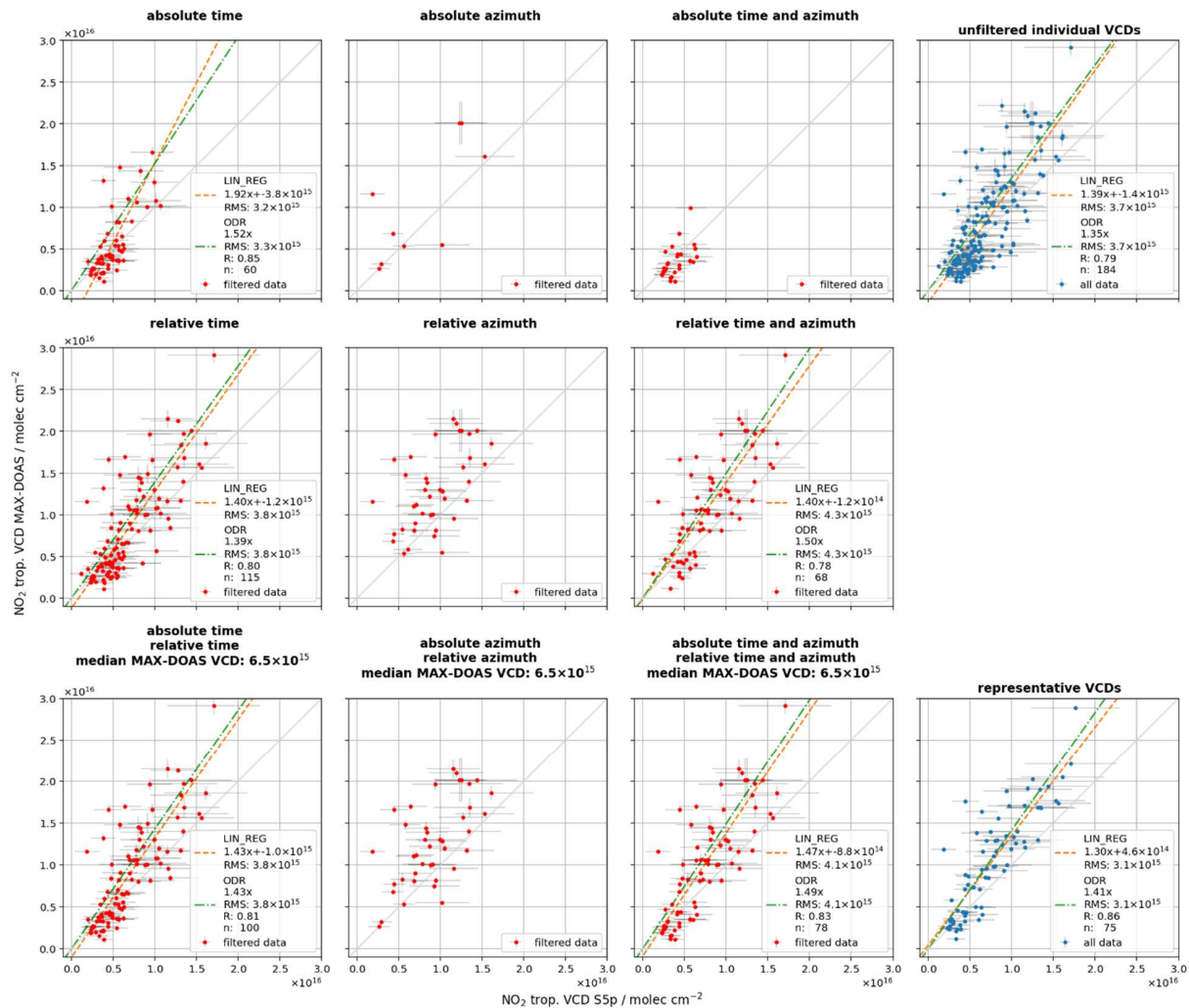
# MAINZ\_1676\_1\_pixel, MAPA, max. 20 km



# MAINZ\_1676\_2\_pixel, MAPA, max. 20 km

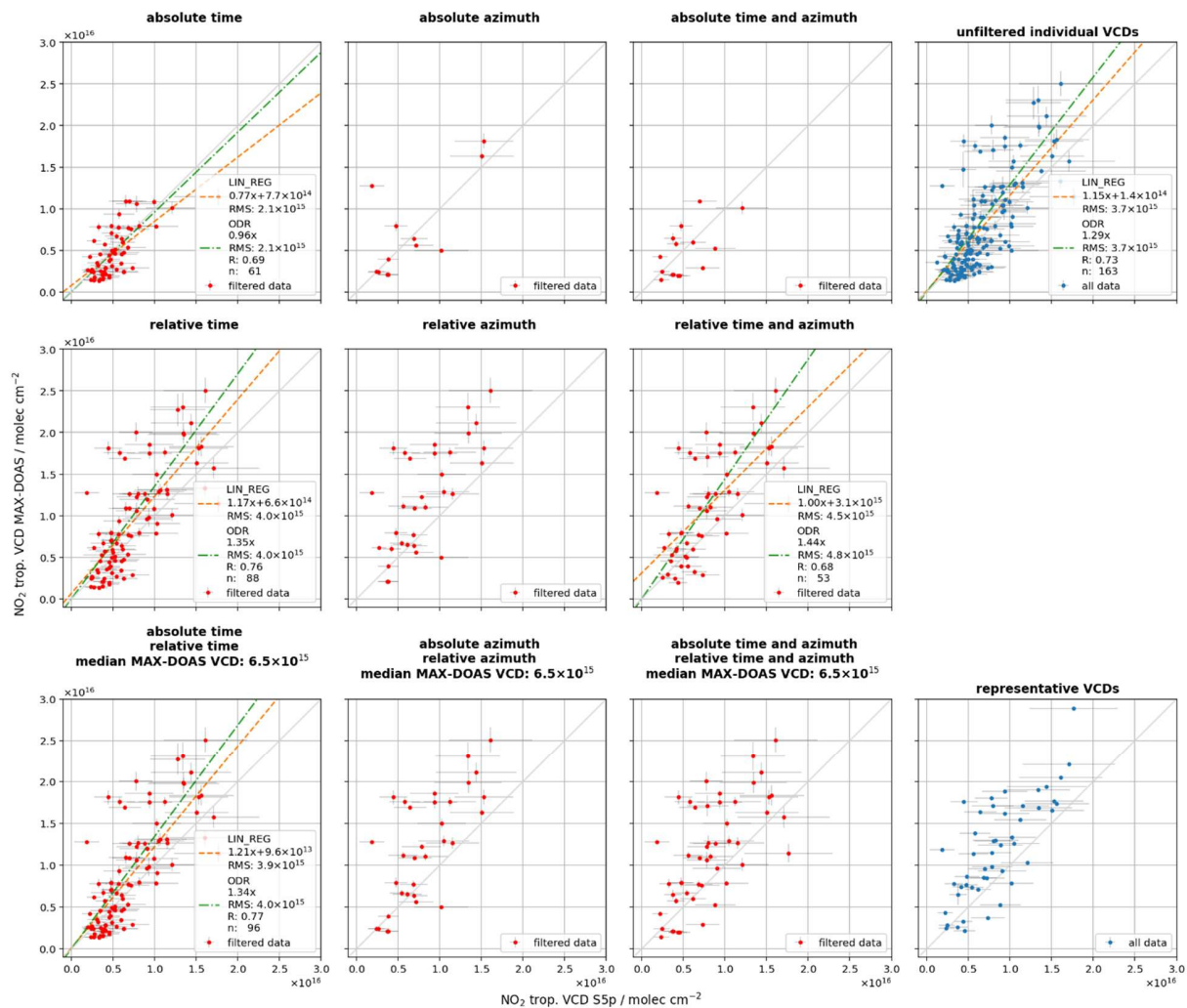


# MAINZ\_1676\_3\_pixel, MAPA, max. 20 km



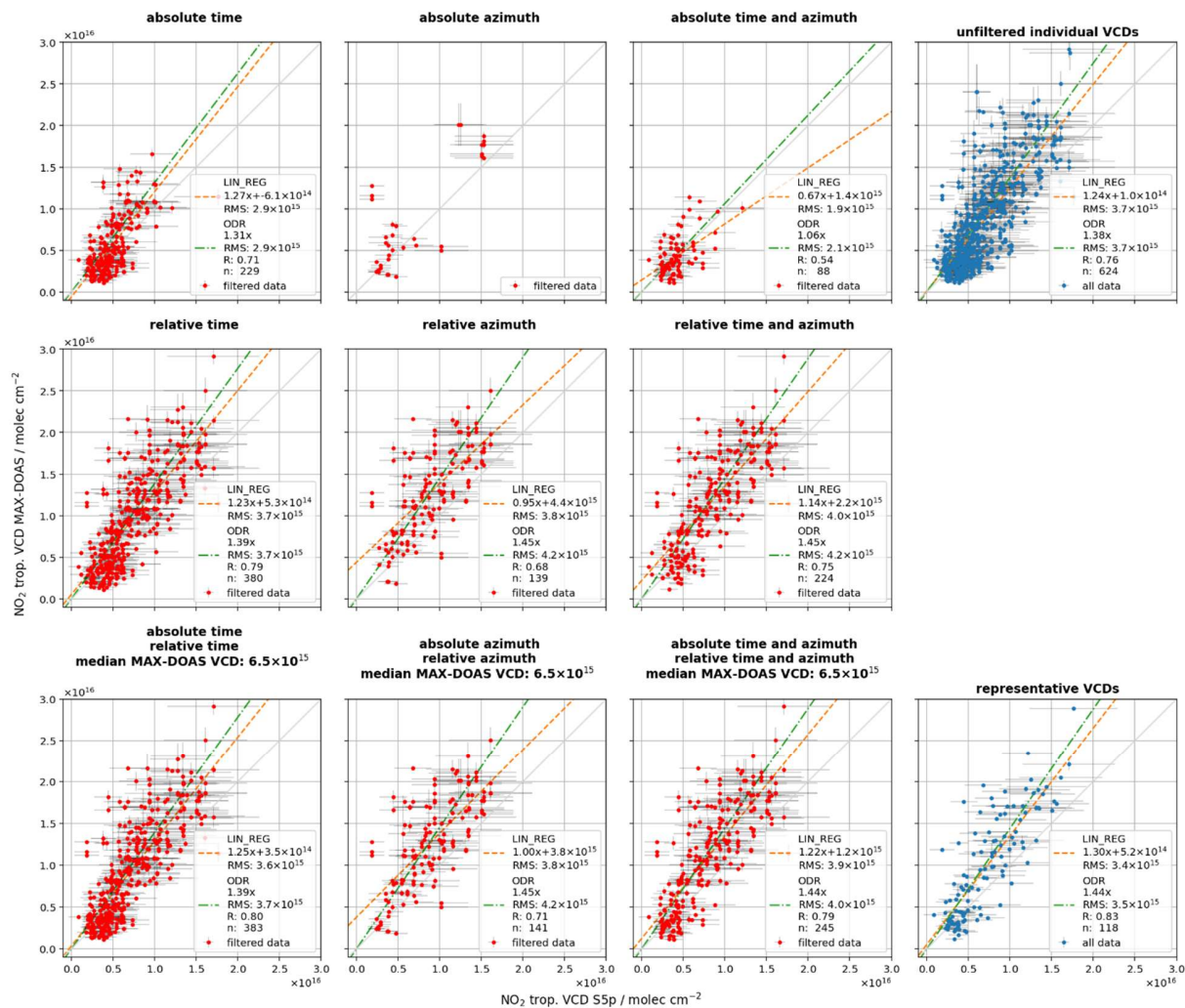


# MAINZ\_1676\_4\_pixel, MAPA, max. 20 km

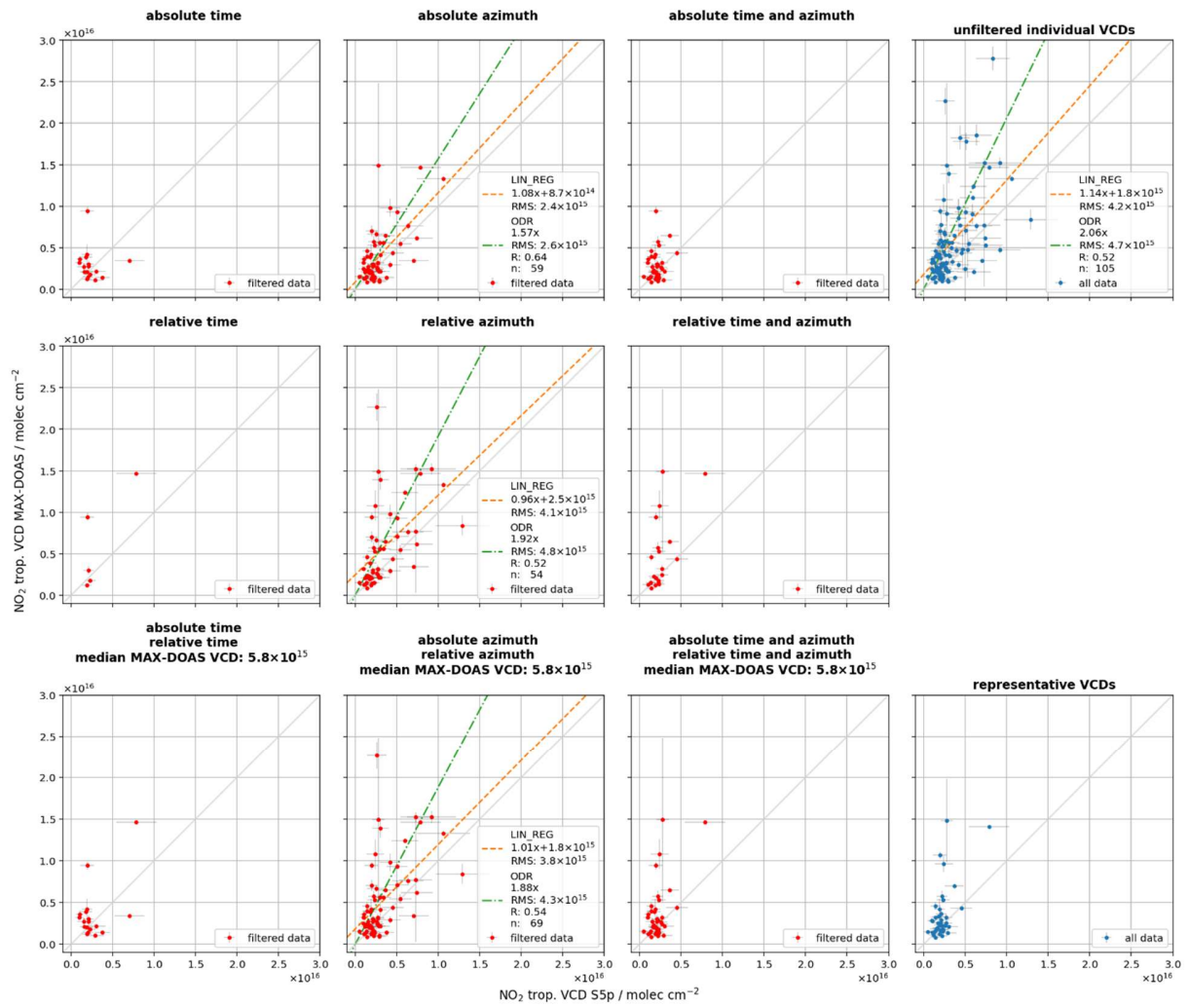




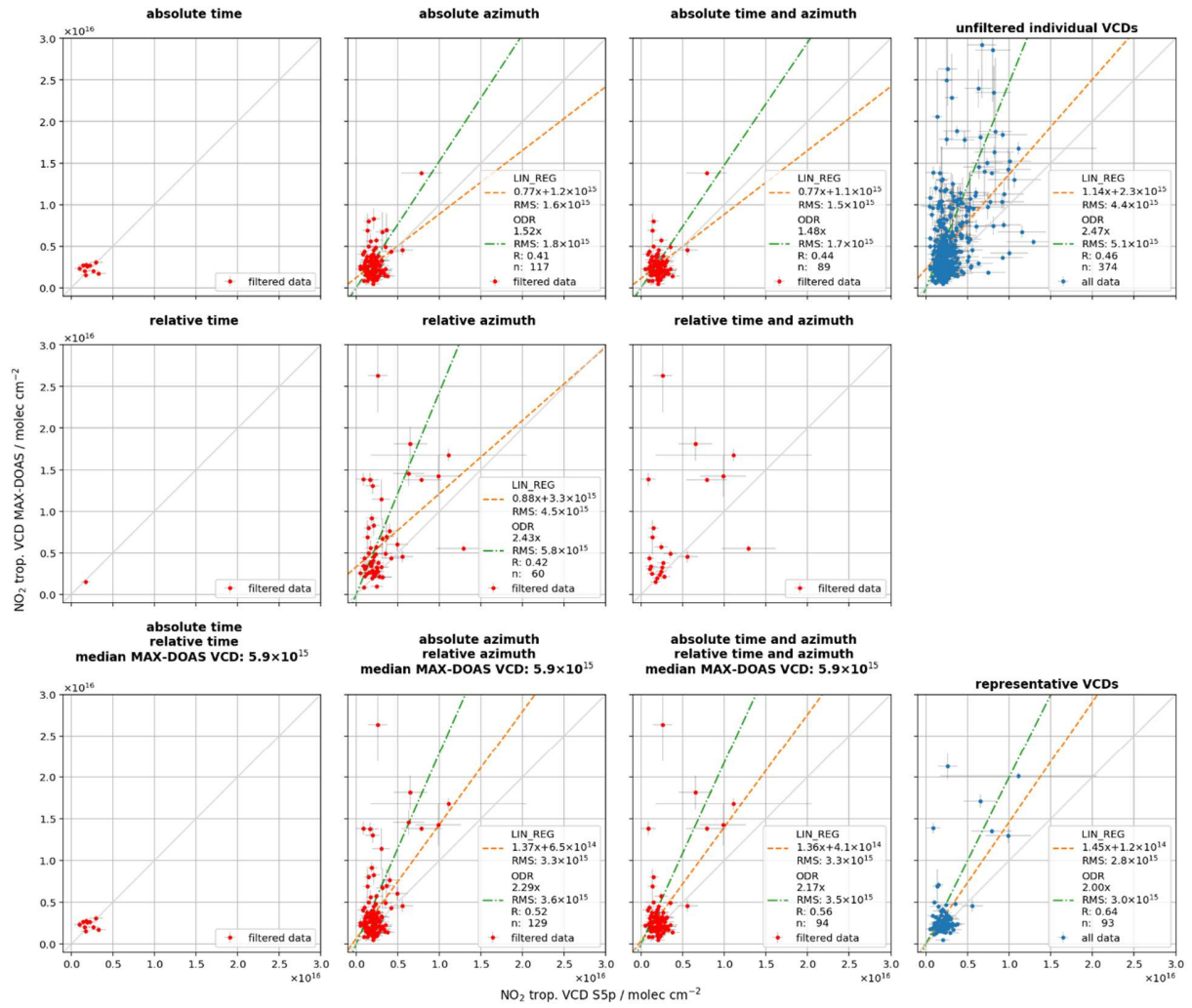
# MAINZ\_1676\_pixel, MAPA, max. 20 km



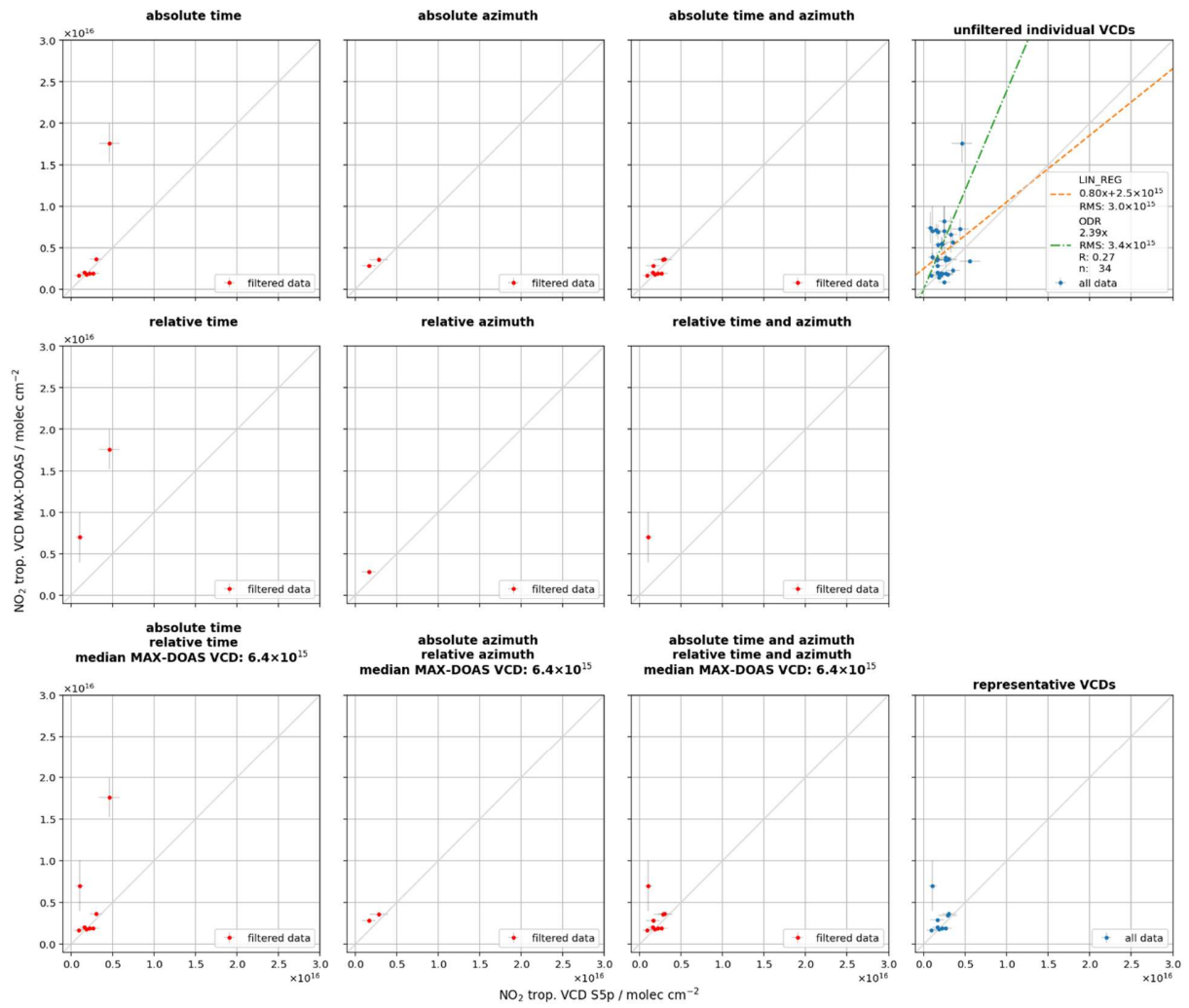
# THESSALONIKI\_1683\_1\_pixel, MAPA, max. 20 km



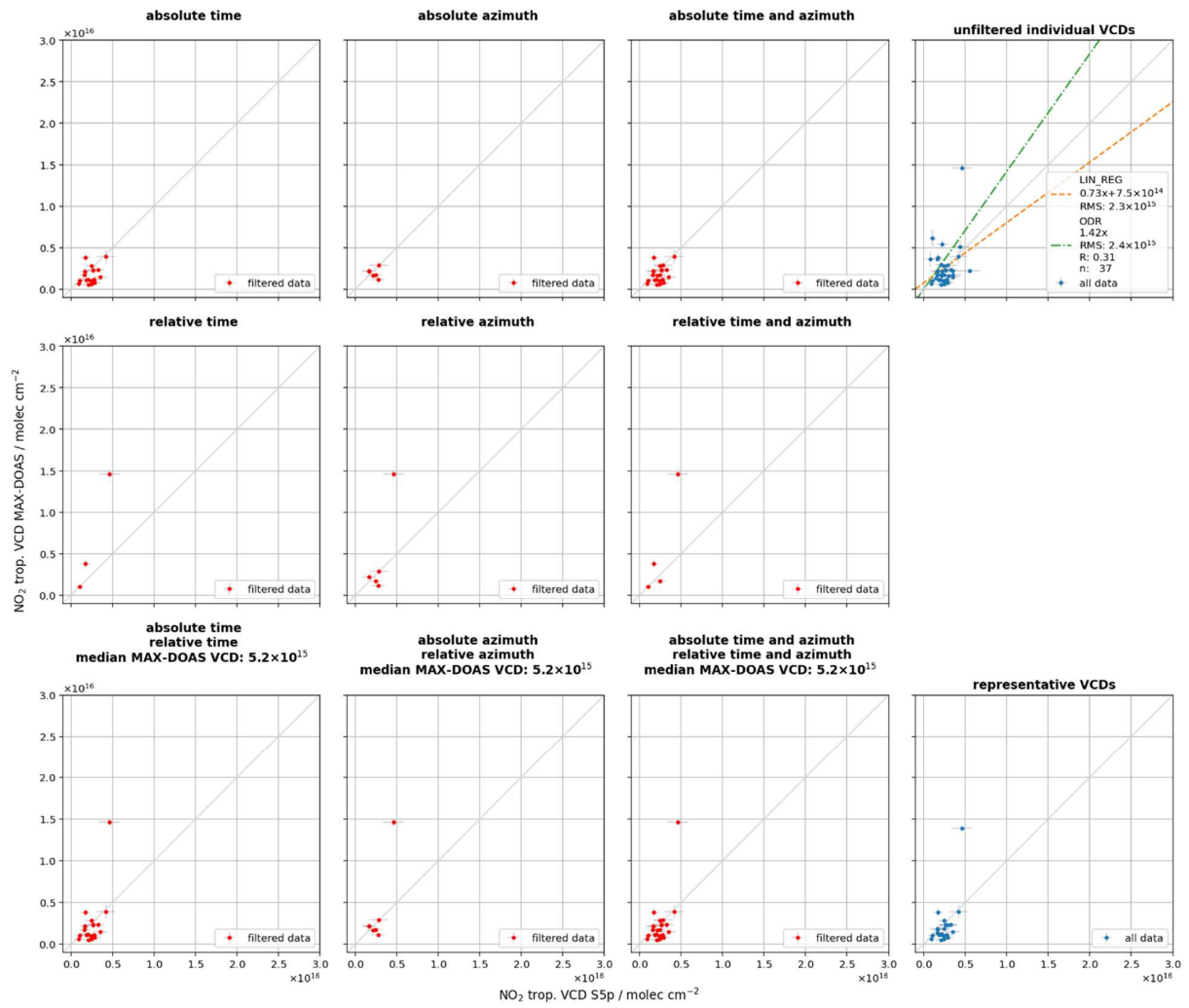
# THESSALONIKI\_1686\_1\_pixel, MAPA, max. 20 km



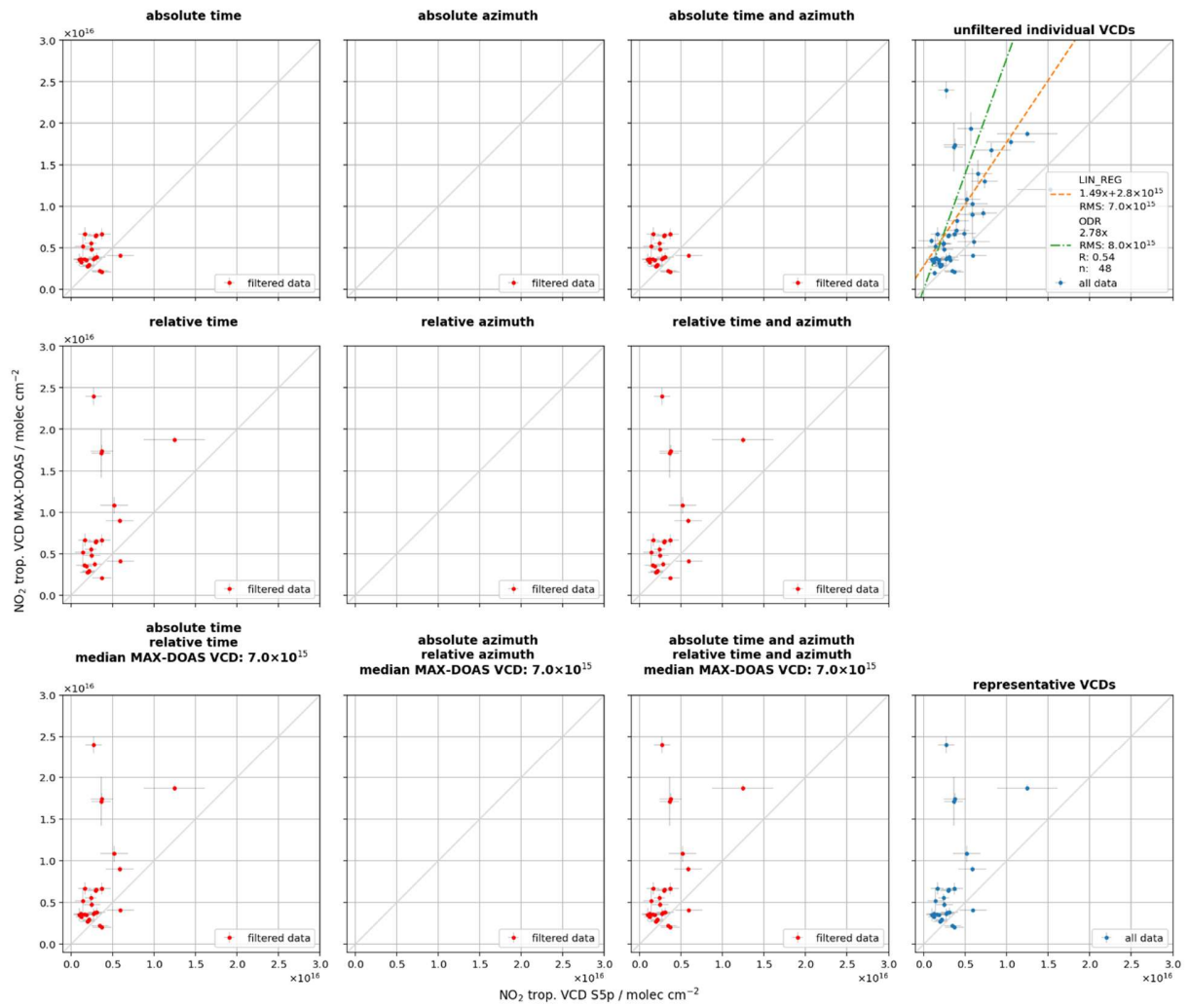
# THESSALONIKI\_1699\_1\_pixel, MAPA, max. 20 km



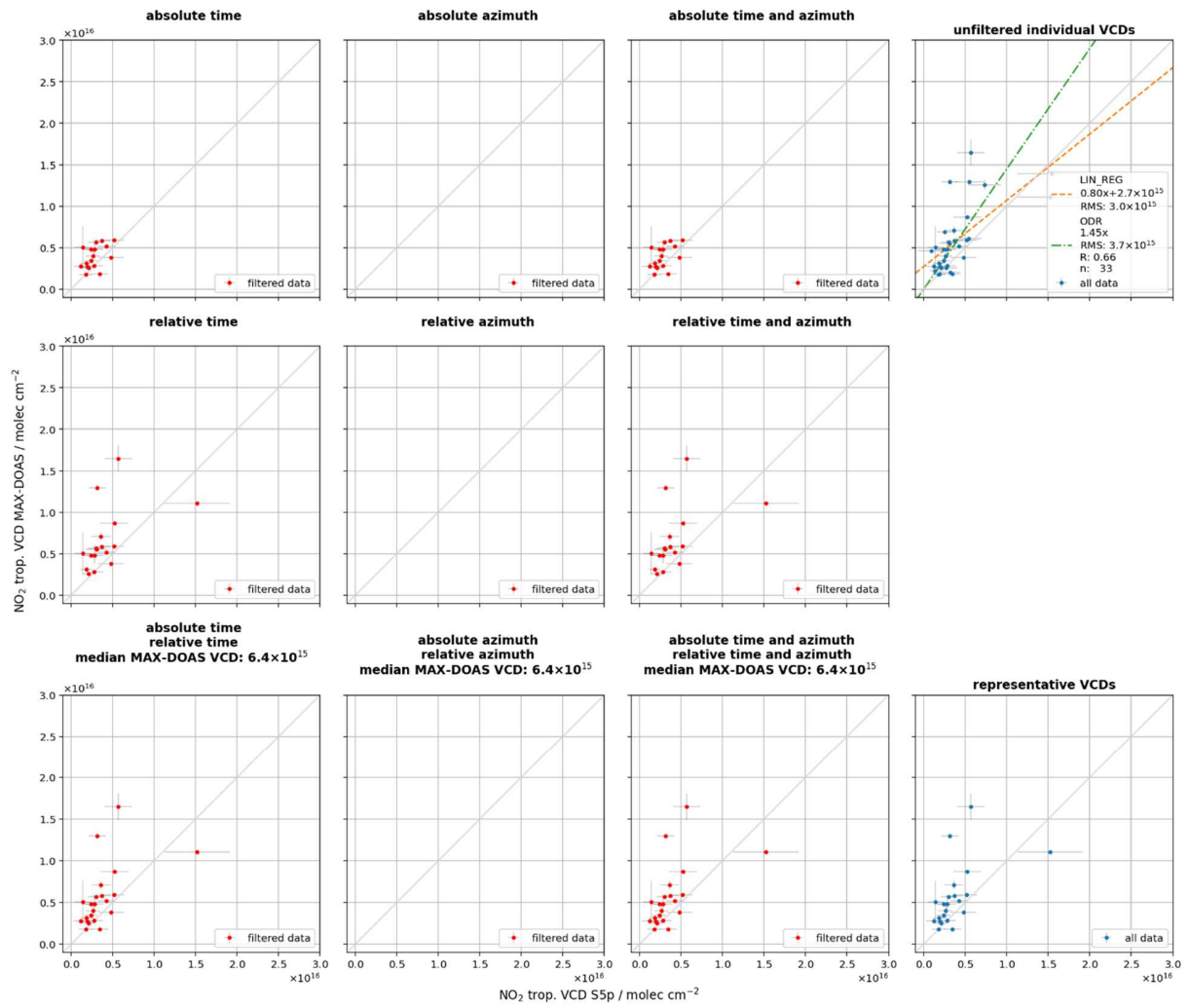
# THESSALONIKI\_1699\_2\_pixel, MAPA, max. 20 km



TORONTO\_1694\_1\_pixel, MAPA, max. 20 km

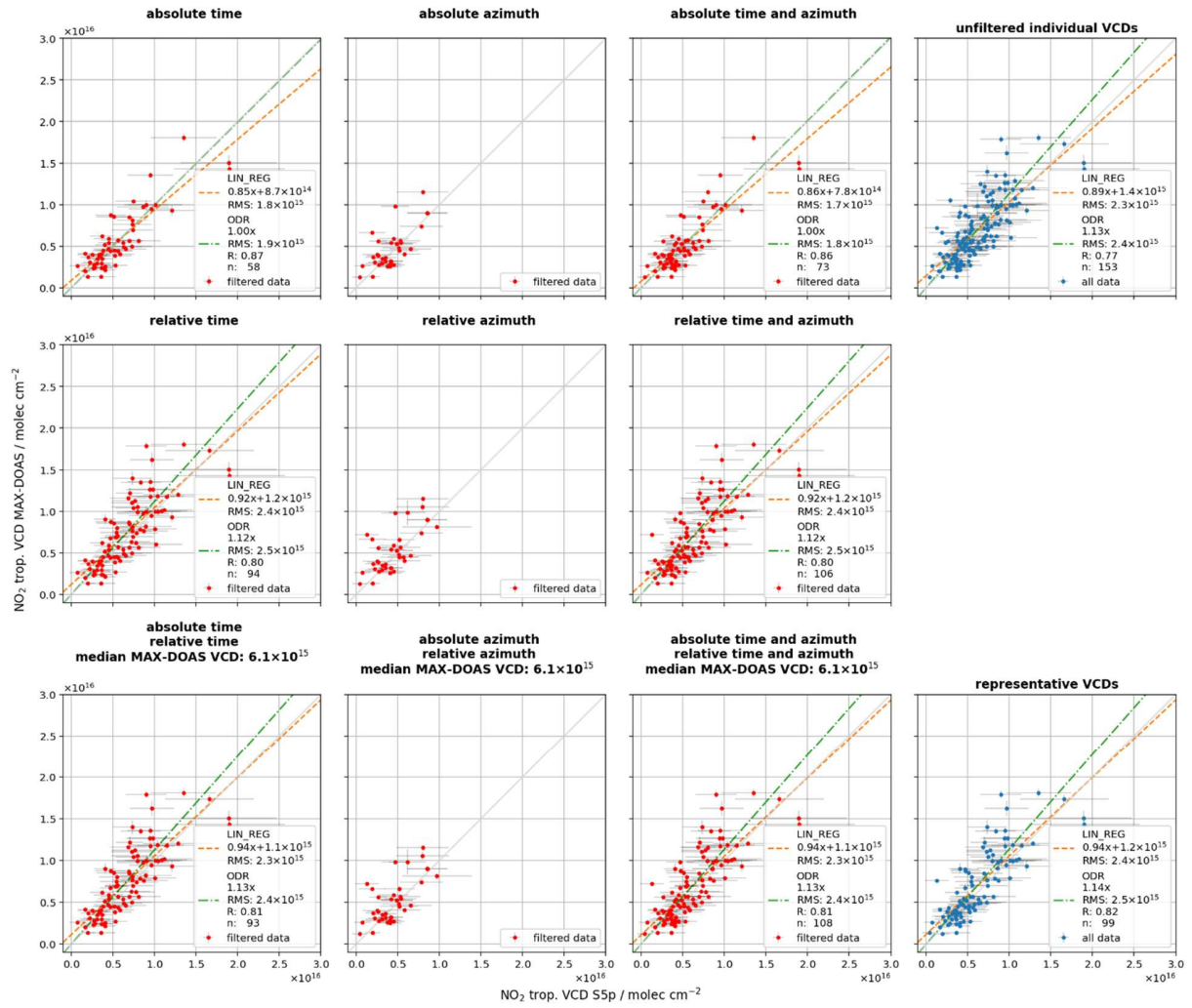


TORONTO\_1694\_2\_pixel, MAPA, max. 20 km

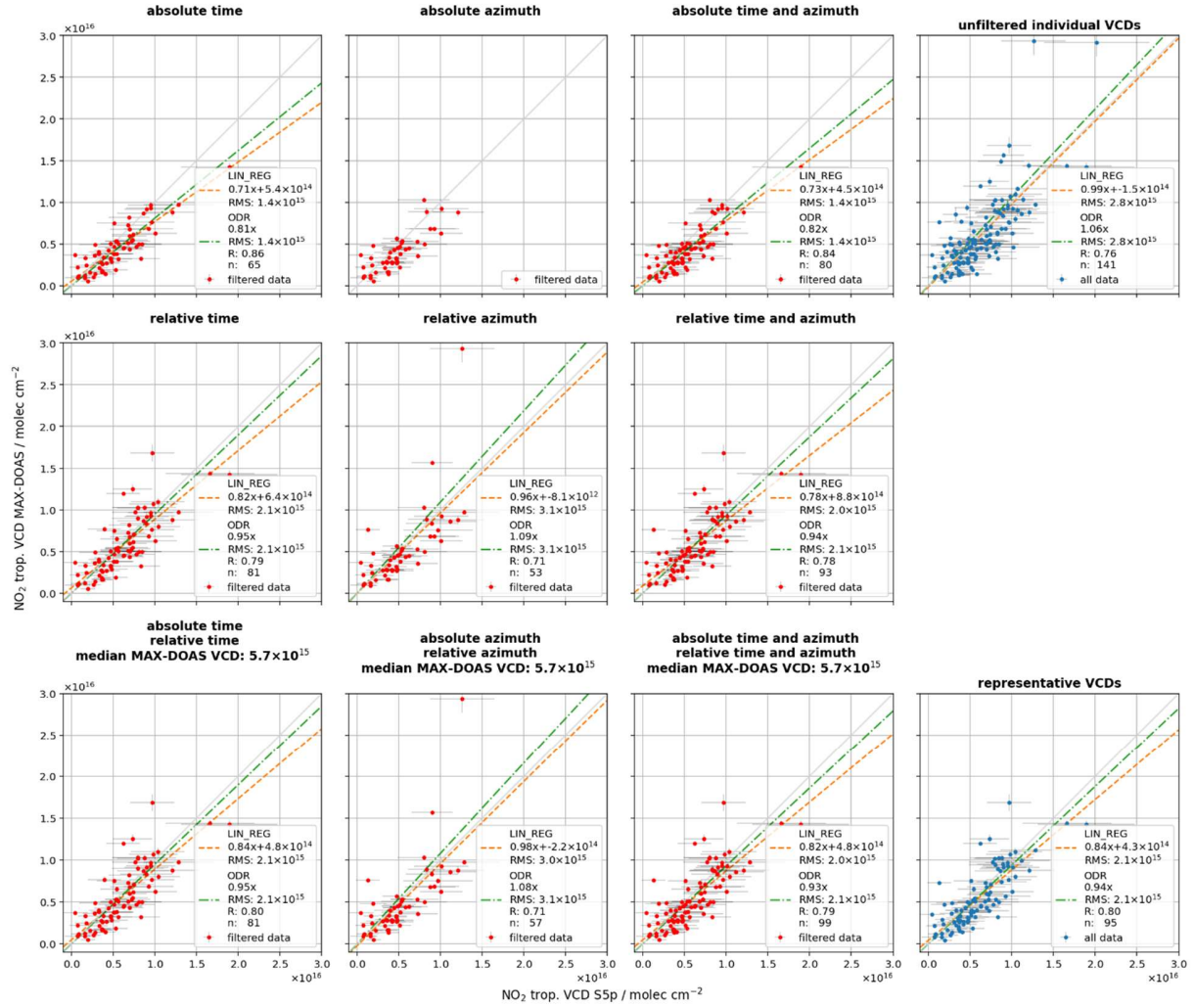




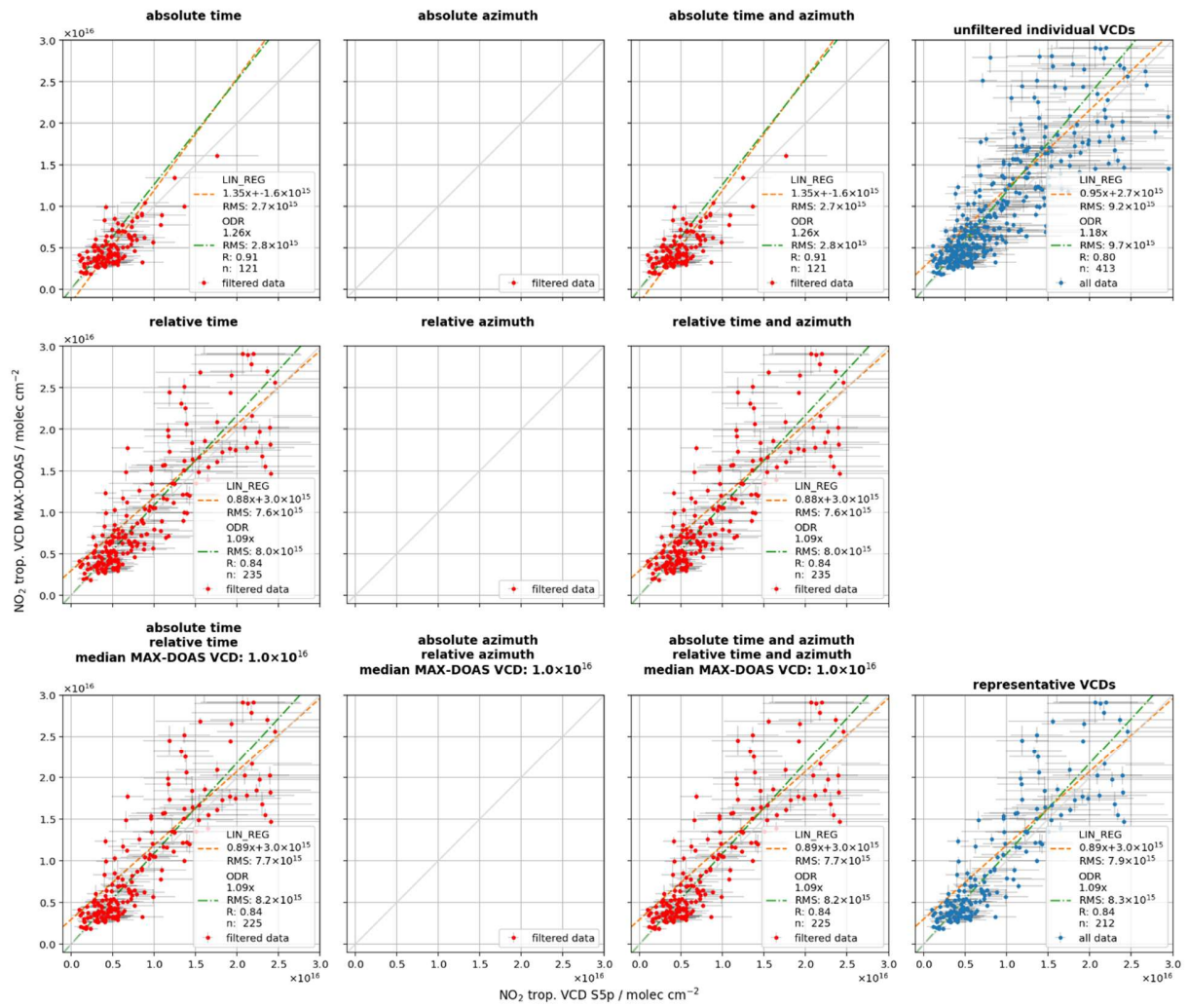
UCCLE\_1670\_1\_pixel, MAPA, max. 20 km



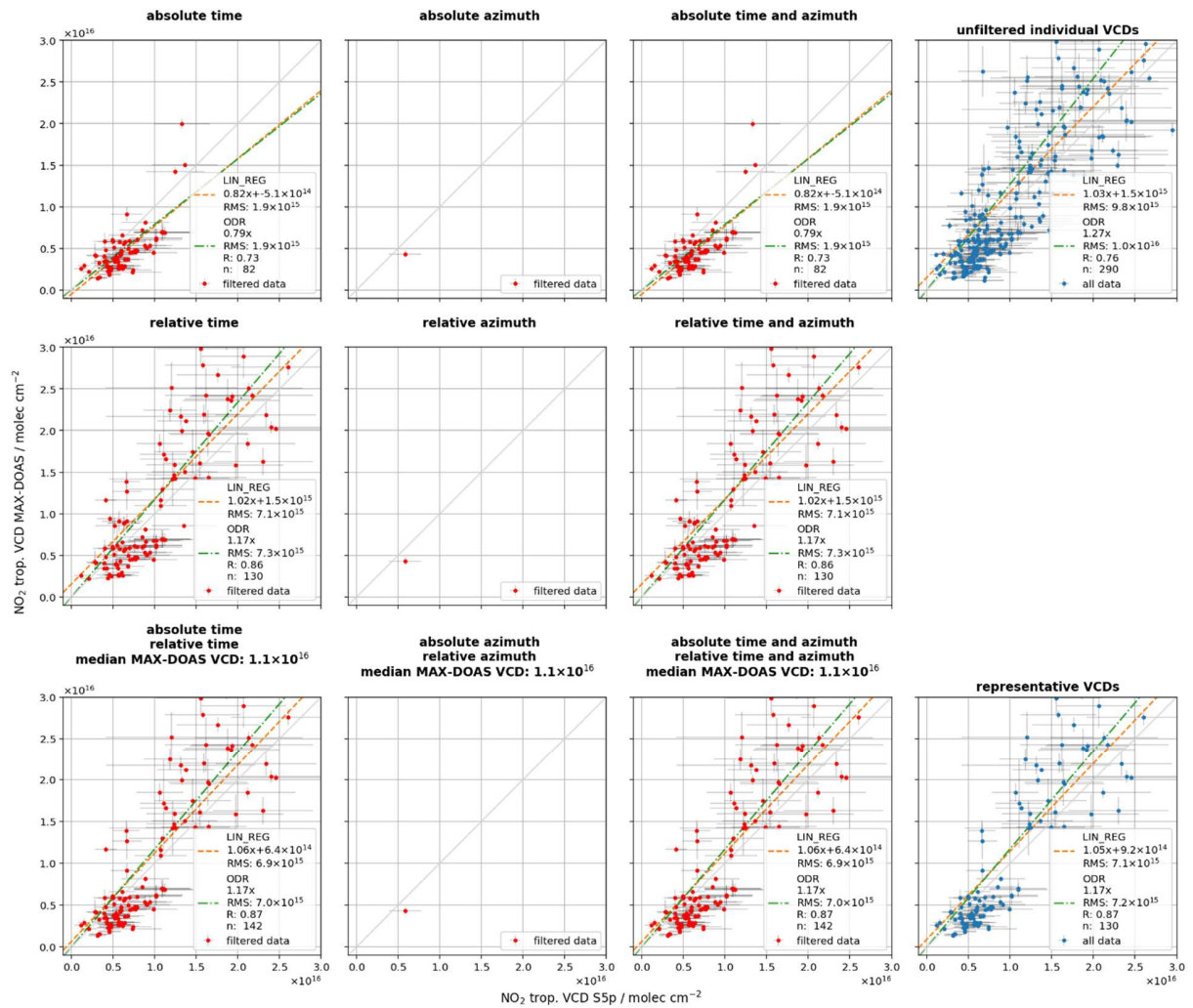
UCCLE\_1670\_2\_pixel, MAPA, max. 20 km



XIANGHE\_1669\_1\_pixel, MAPA, max. 20 km

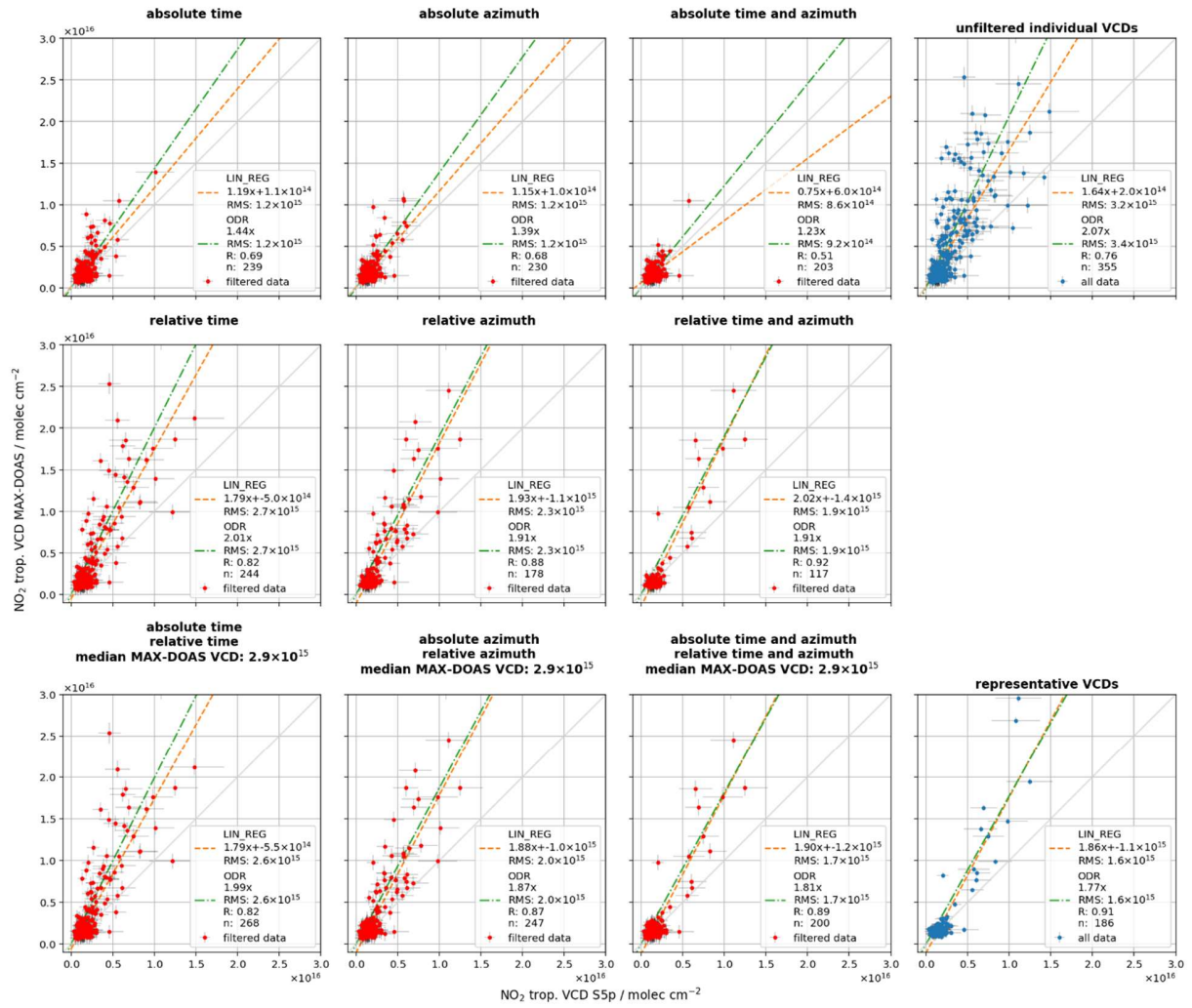


# XIANGHE\_1669\_2\_pixel, MAPA, max. 20 km



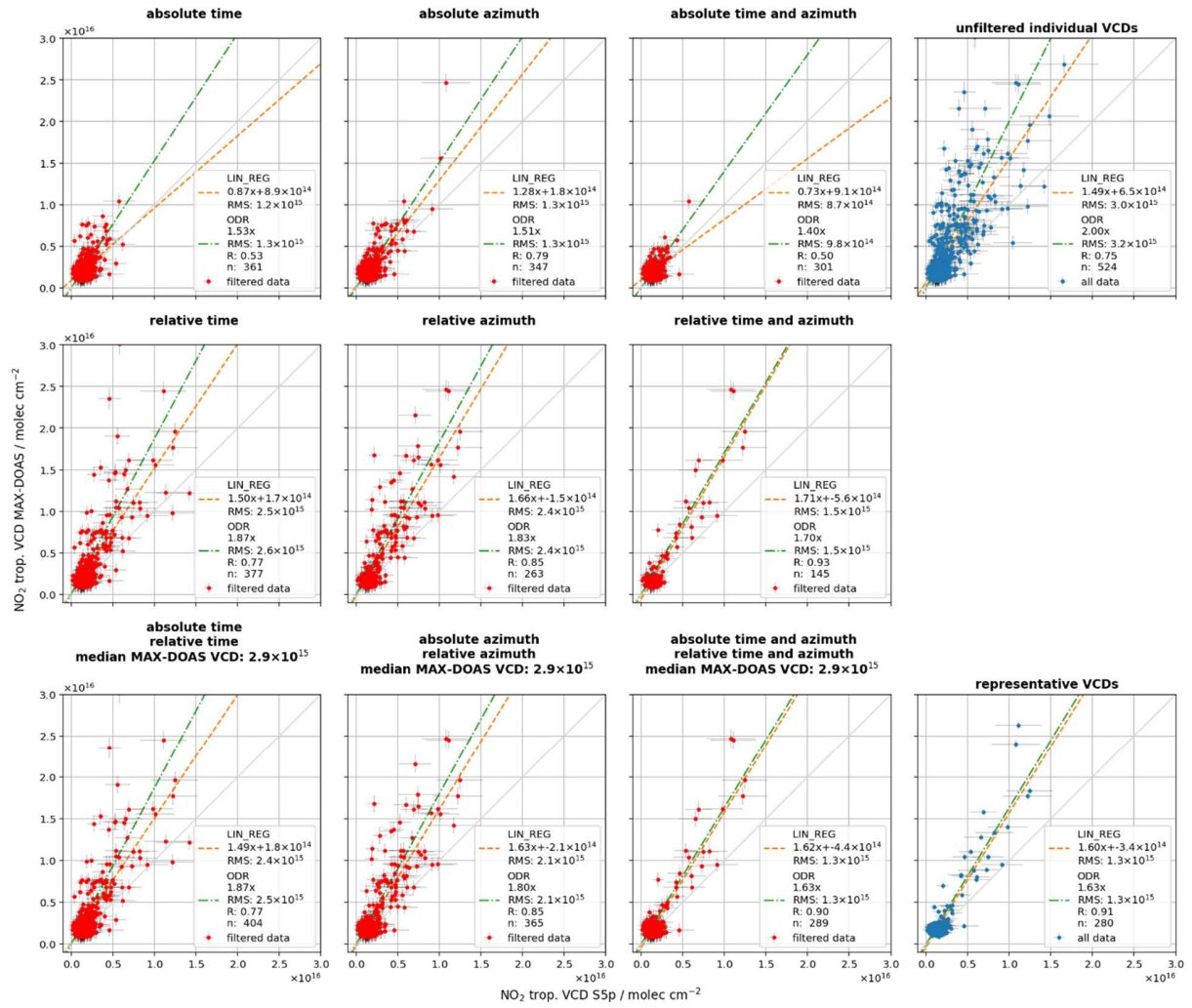
## Scatterplots MMF

# ATHENS\_1674\_1\_pixel, MMF, max. 20 km



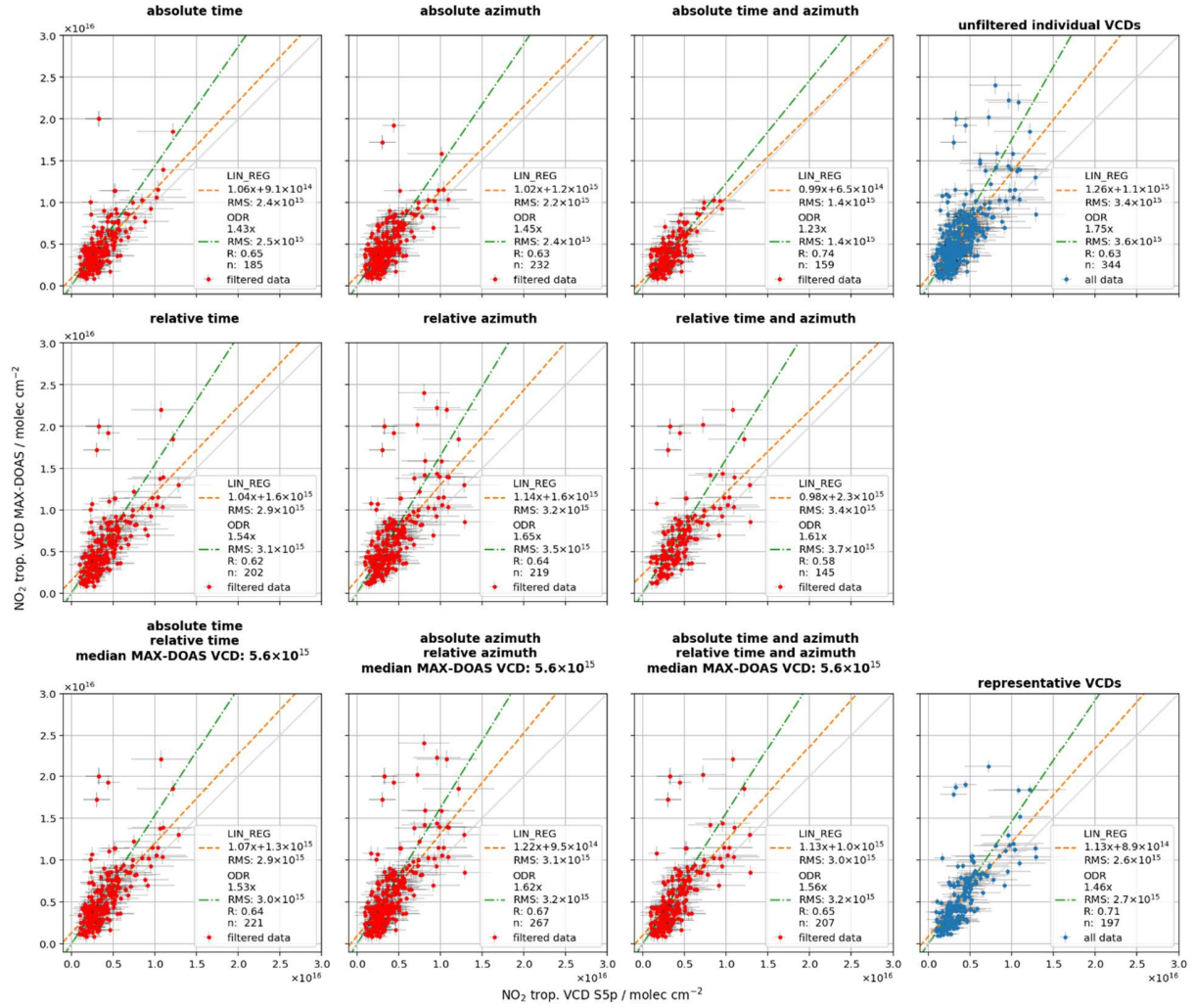


# ATHENSuv\_1674\_1\_pixel, MMF, max. 20 km

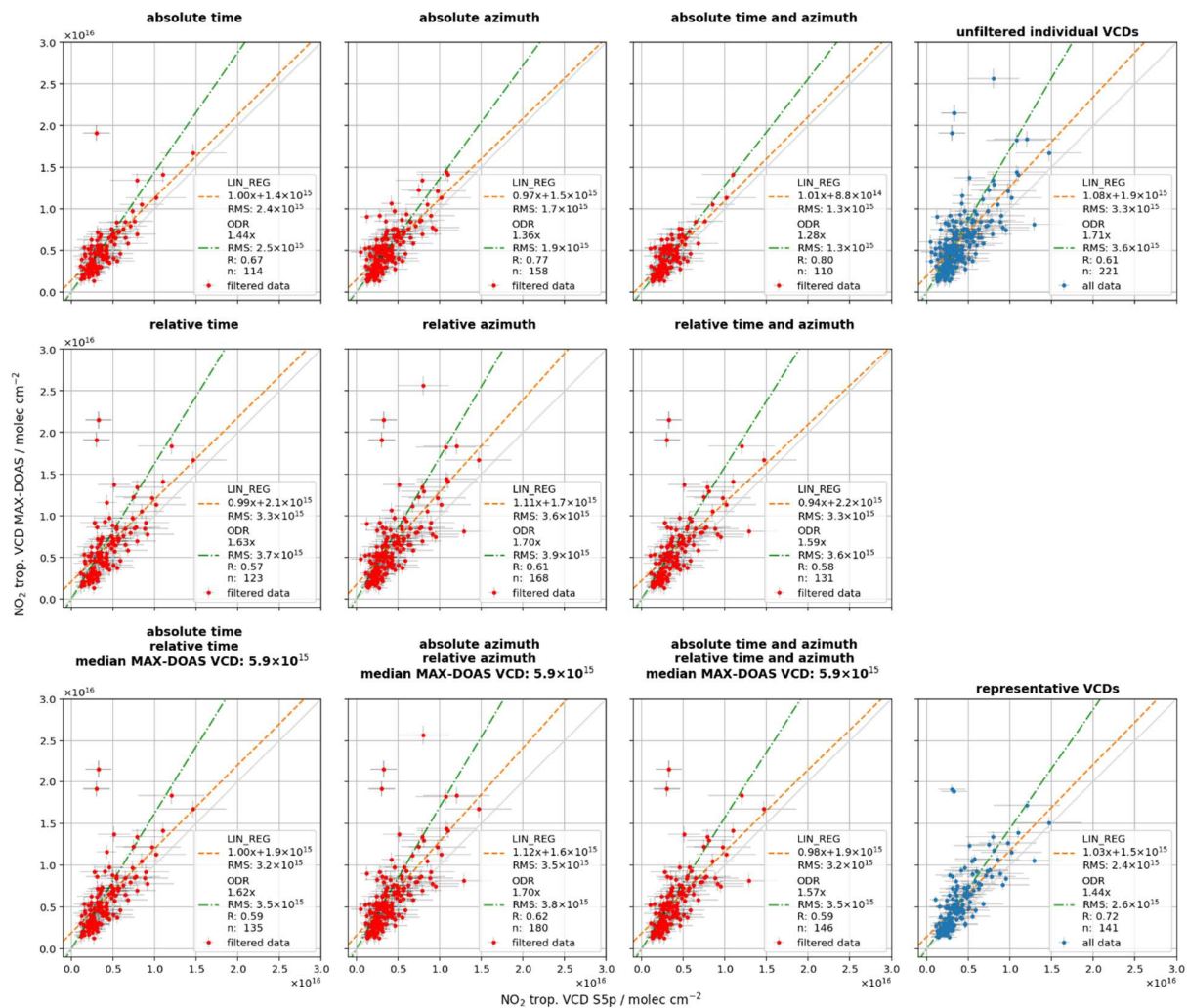




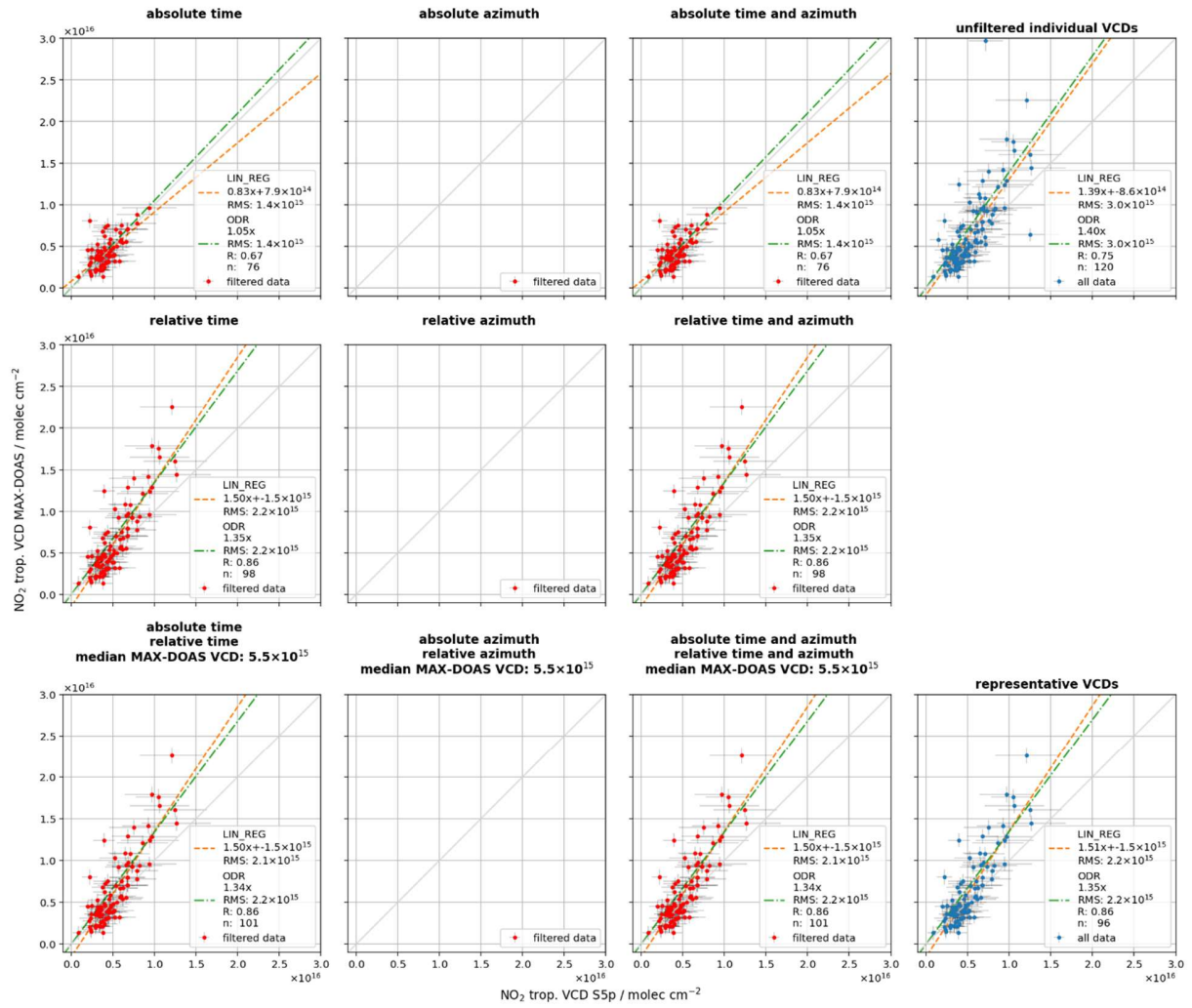
# BREMEN\_1673\_1\_pixel, MMF, max. 20 km



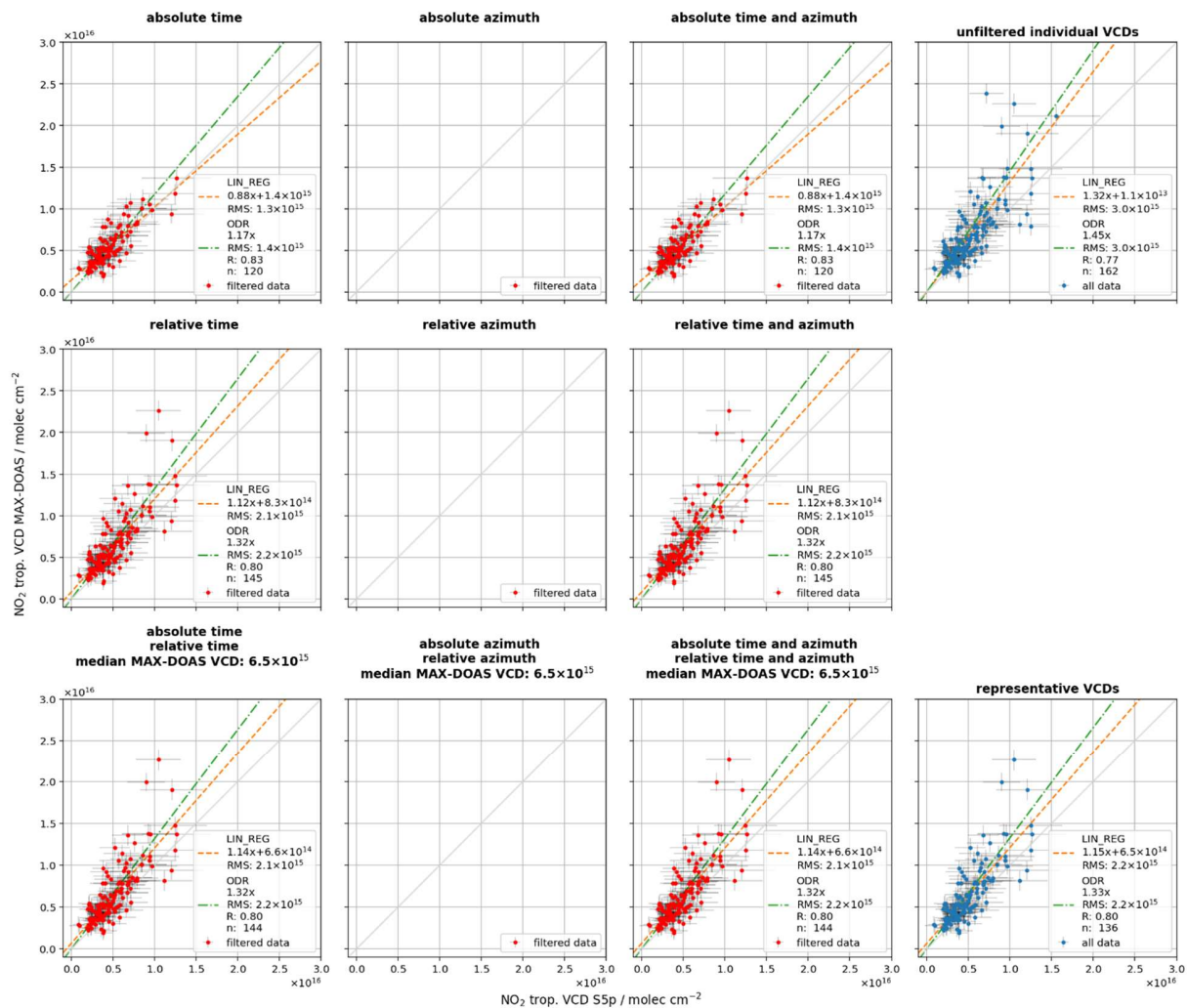
# BREMEN\_1673\_2\_pixel, MMF, max. 20 km



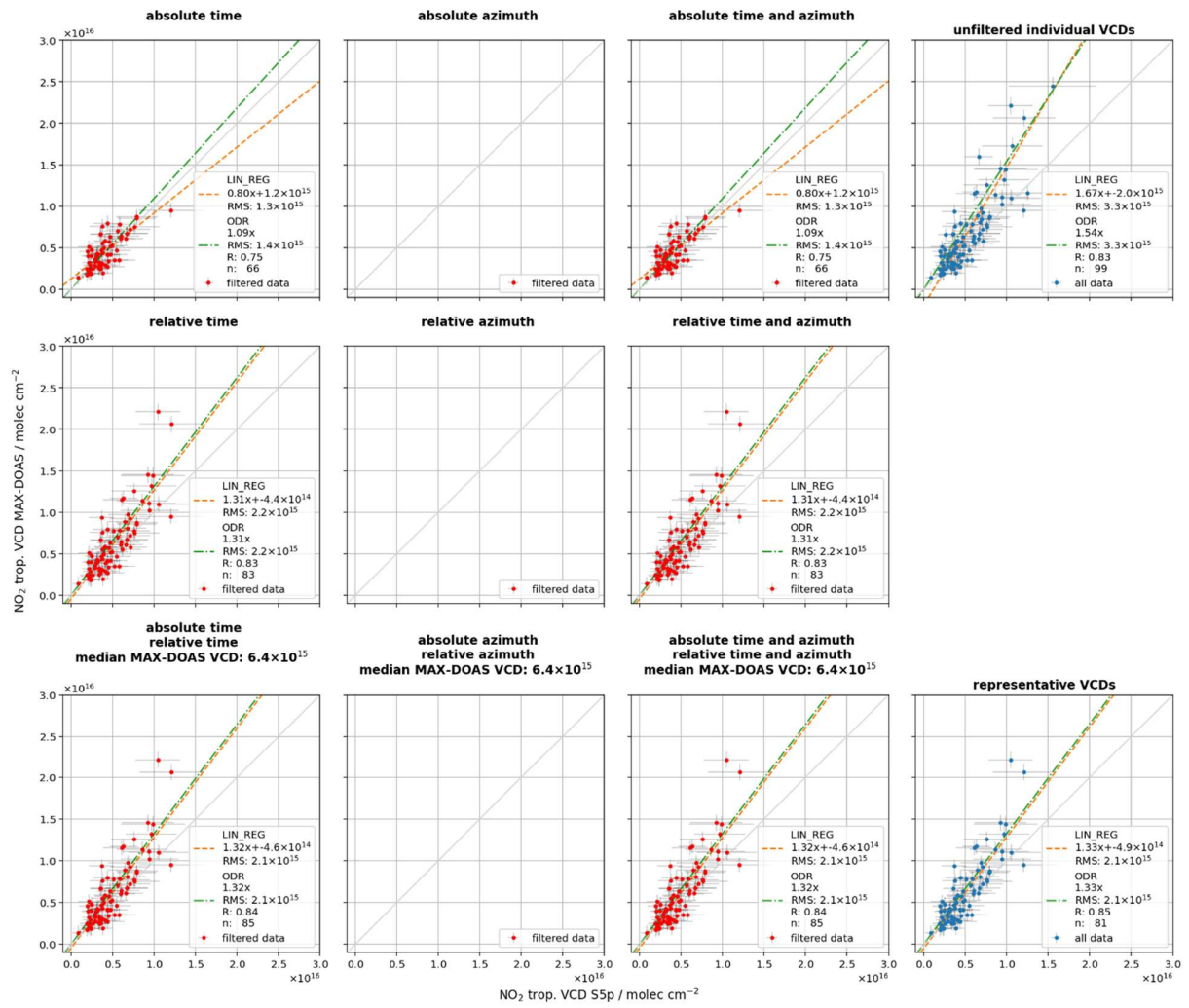
CABAUW\_1690\_1\_pixel, MMF, max. 20 km



CABAUW\_1693\_1\_pixel, MMF, max. 20 km

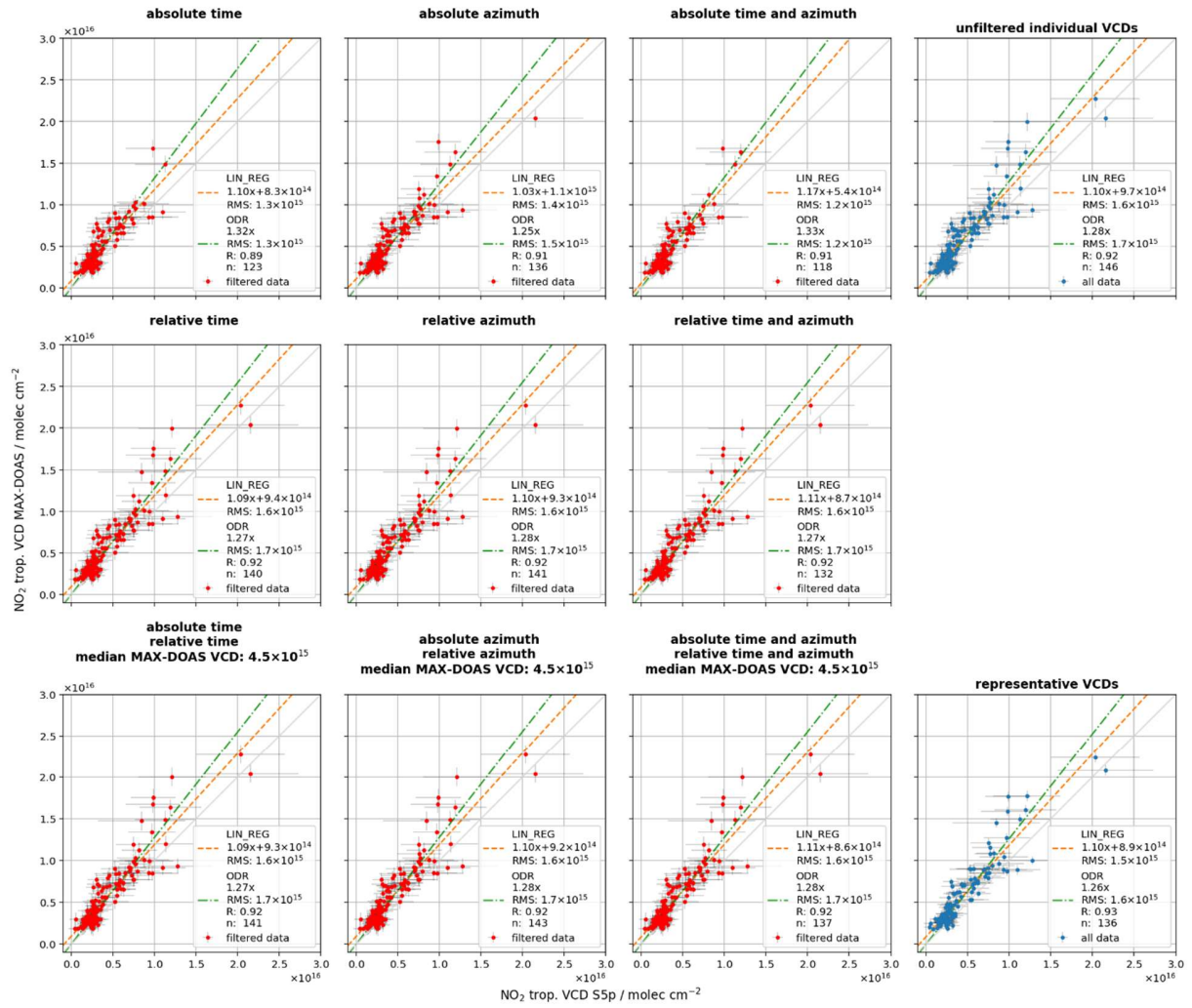


CABAUW\_1693\_2\_pixel, MMF, max. 20 km

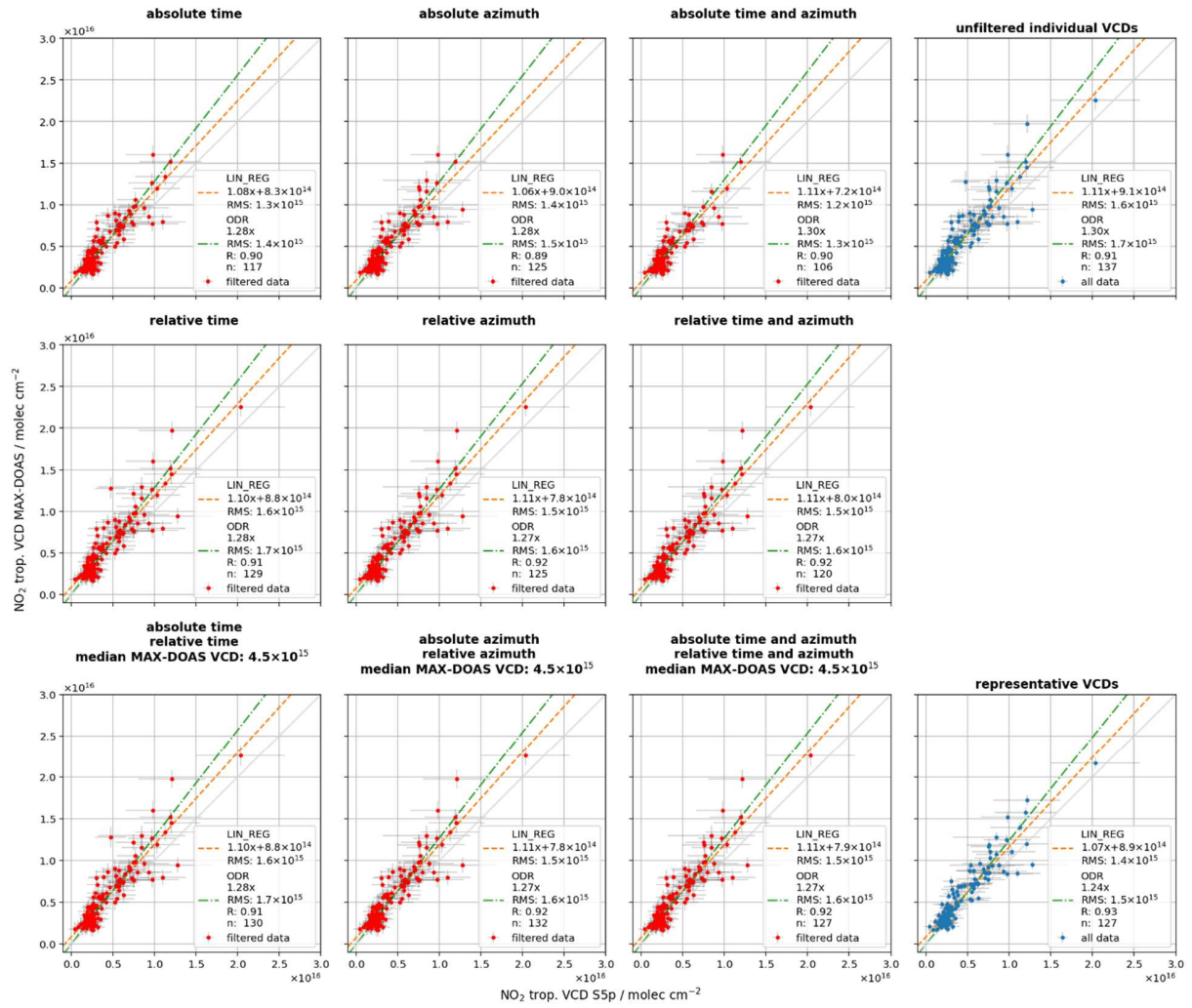




CAPOFIUME\_1695\_1\_pixel, MMF, max. 20 km

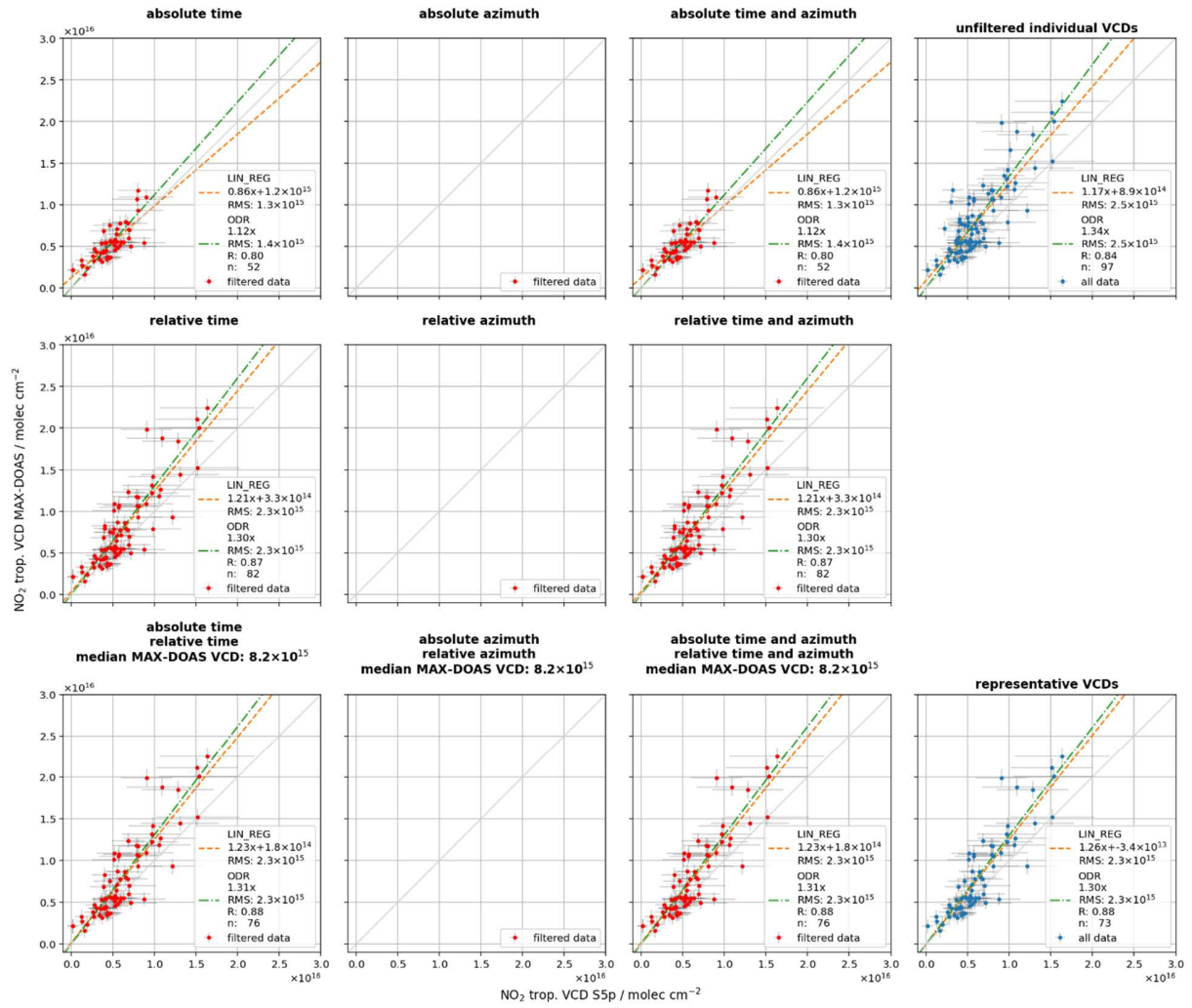


CAPOFIUME\_1695\_2\_pixel, MMF, max. 20 km

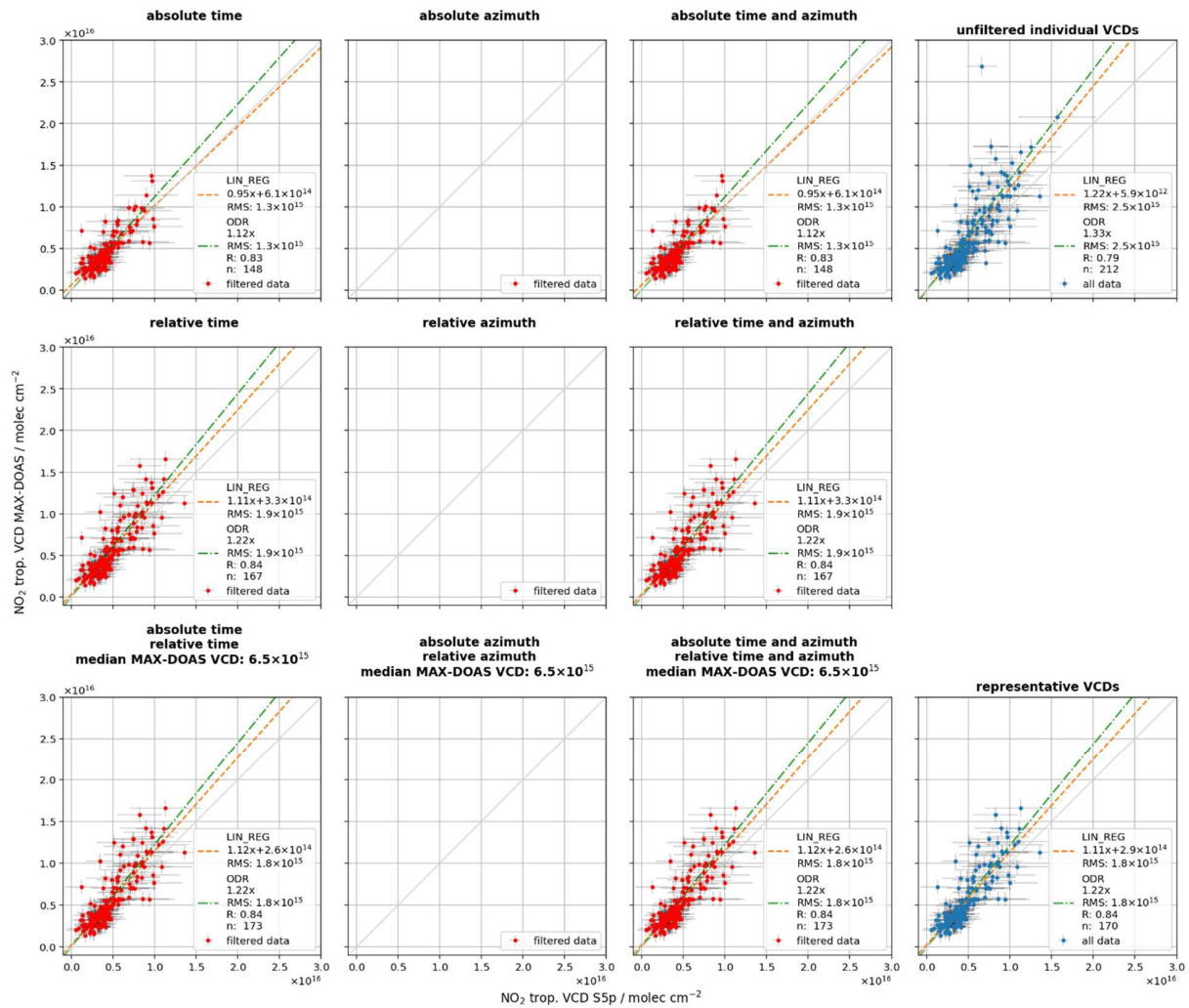




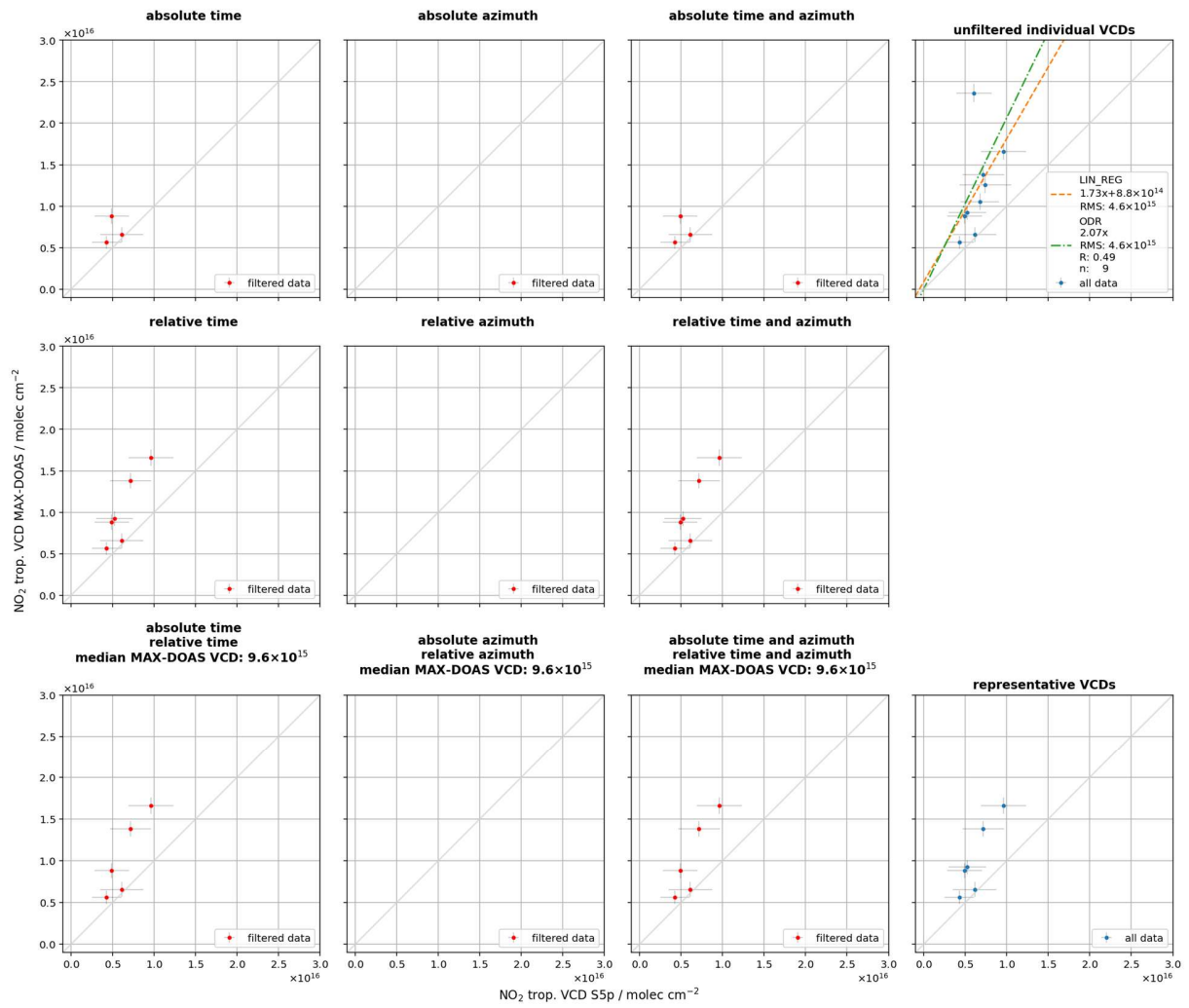
DEBILT\_1675\_1\_pixel, MMF, max. 20 km



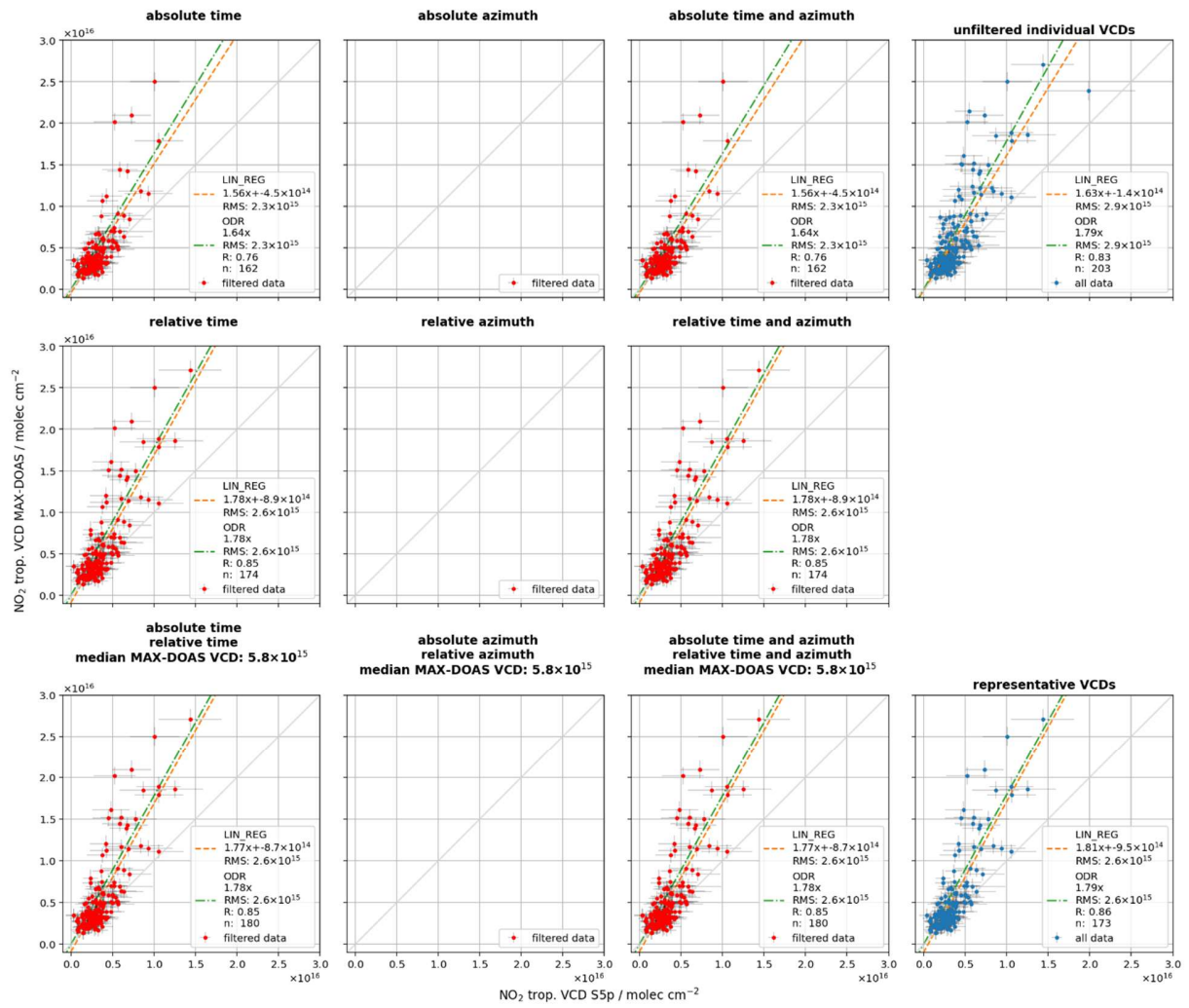
DEBILT\_1684\_1\_pixel, MMF, max. 20 km



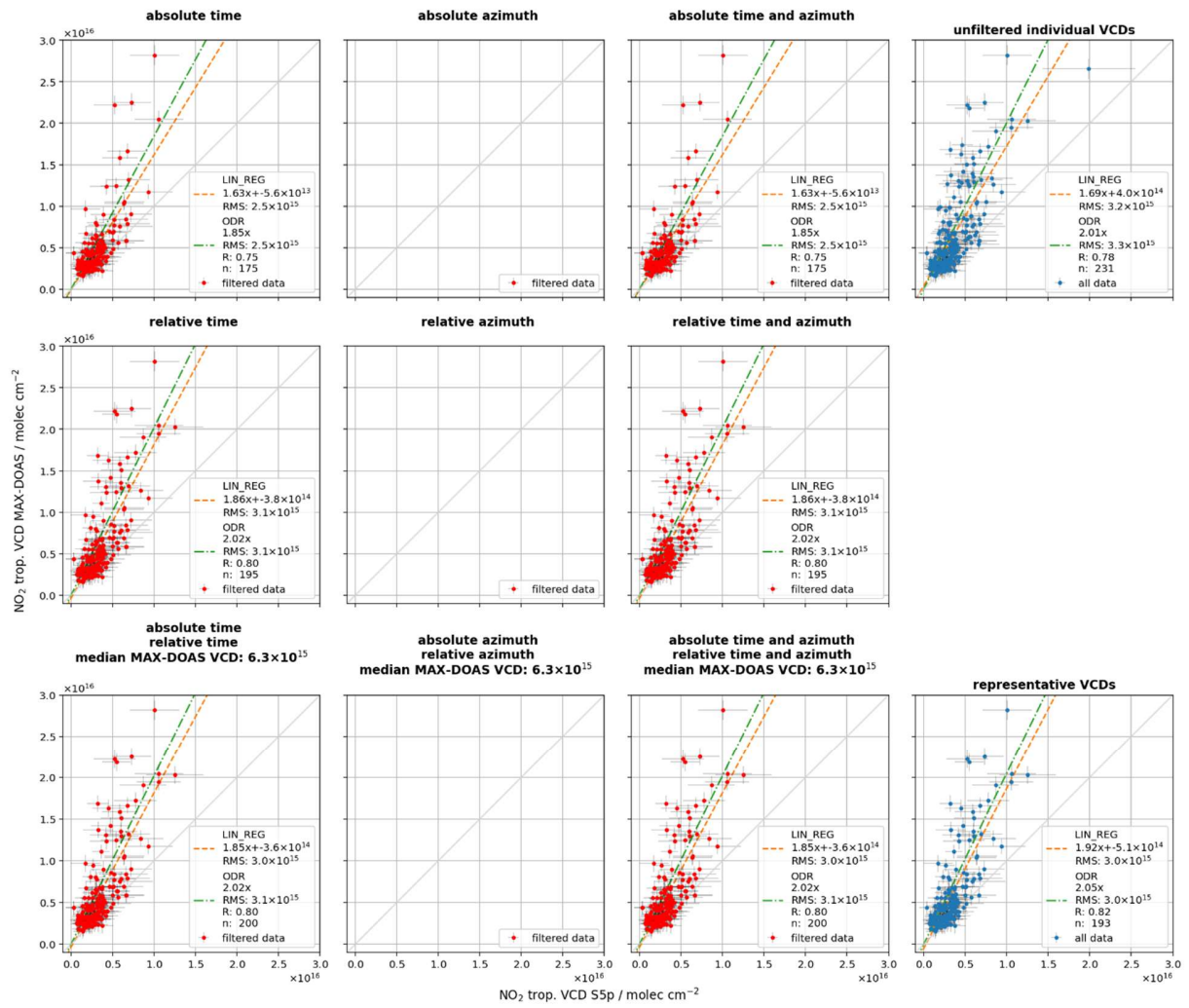
DEBILT\_1688\_1\_pixel, MMF, max. 20 km



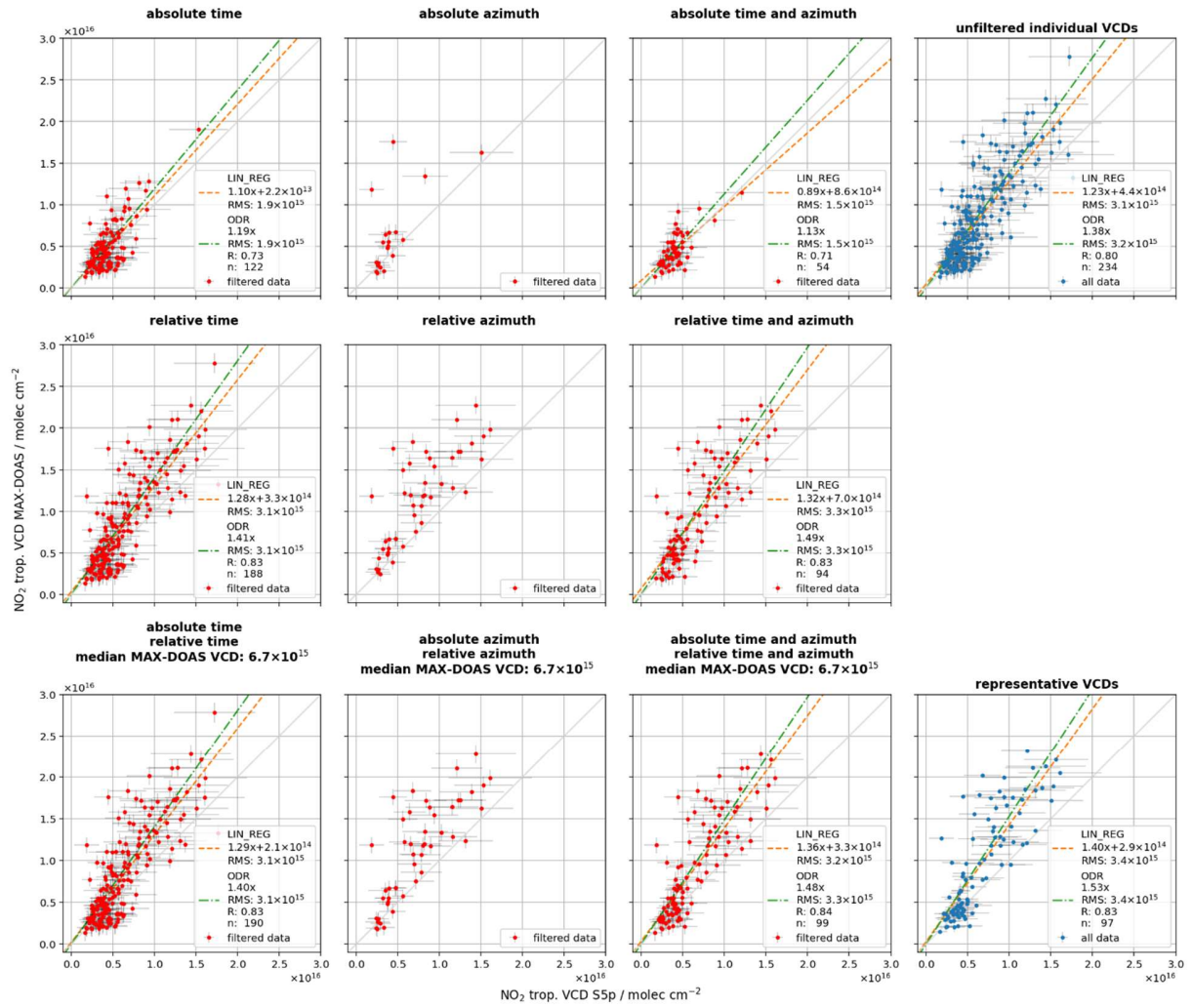
# HEIDELBERG\_1679\_1\_pixel, MMF, max. 20 km



# HEIDELBERG\_1679\_2\_pixel, MMF, max. 20 km

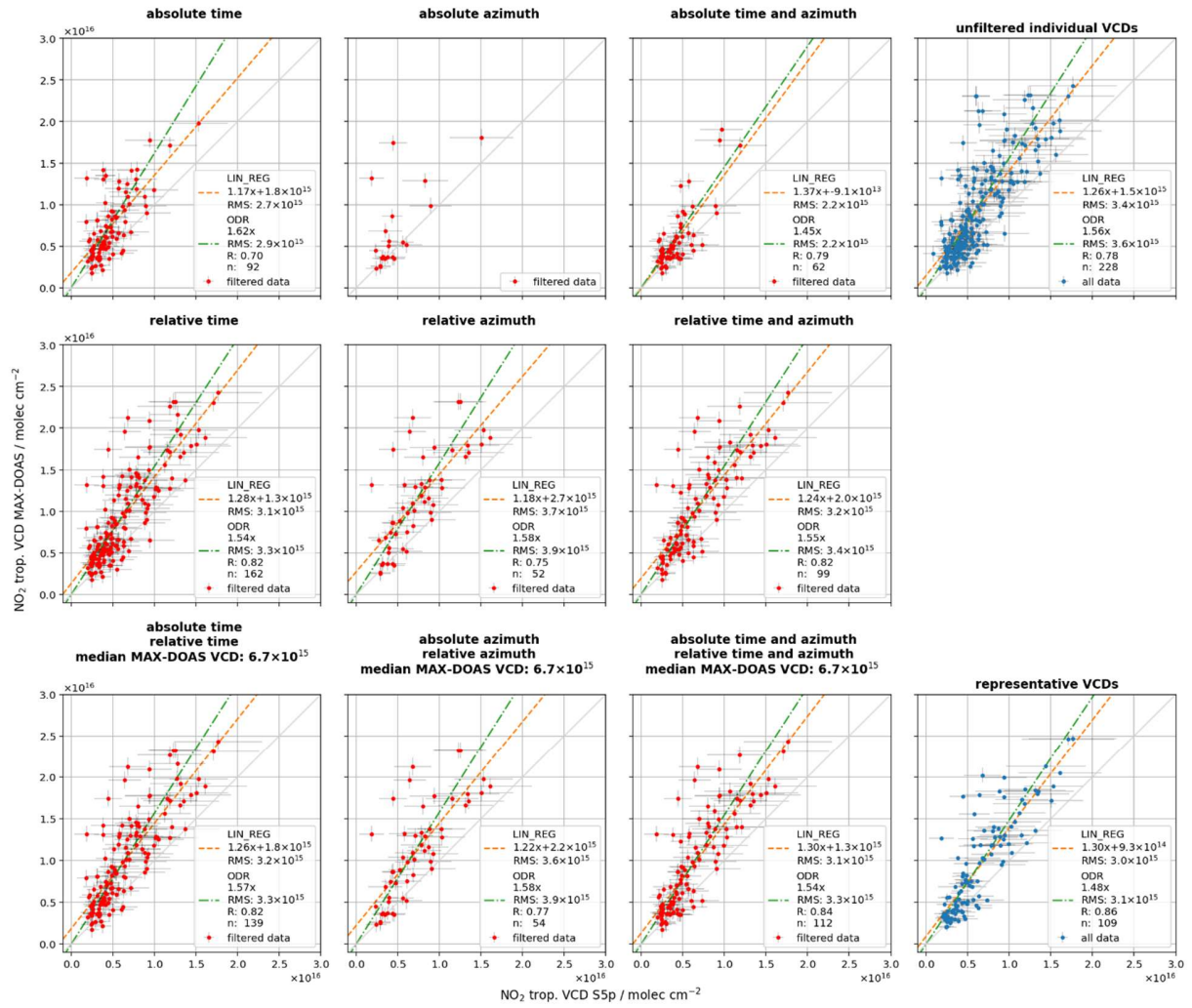


# MAINZ\_1676\_1\_pixel, MMF, max. 20 km



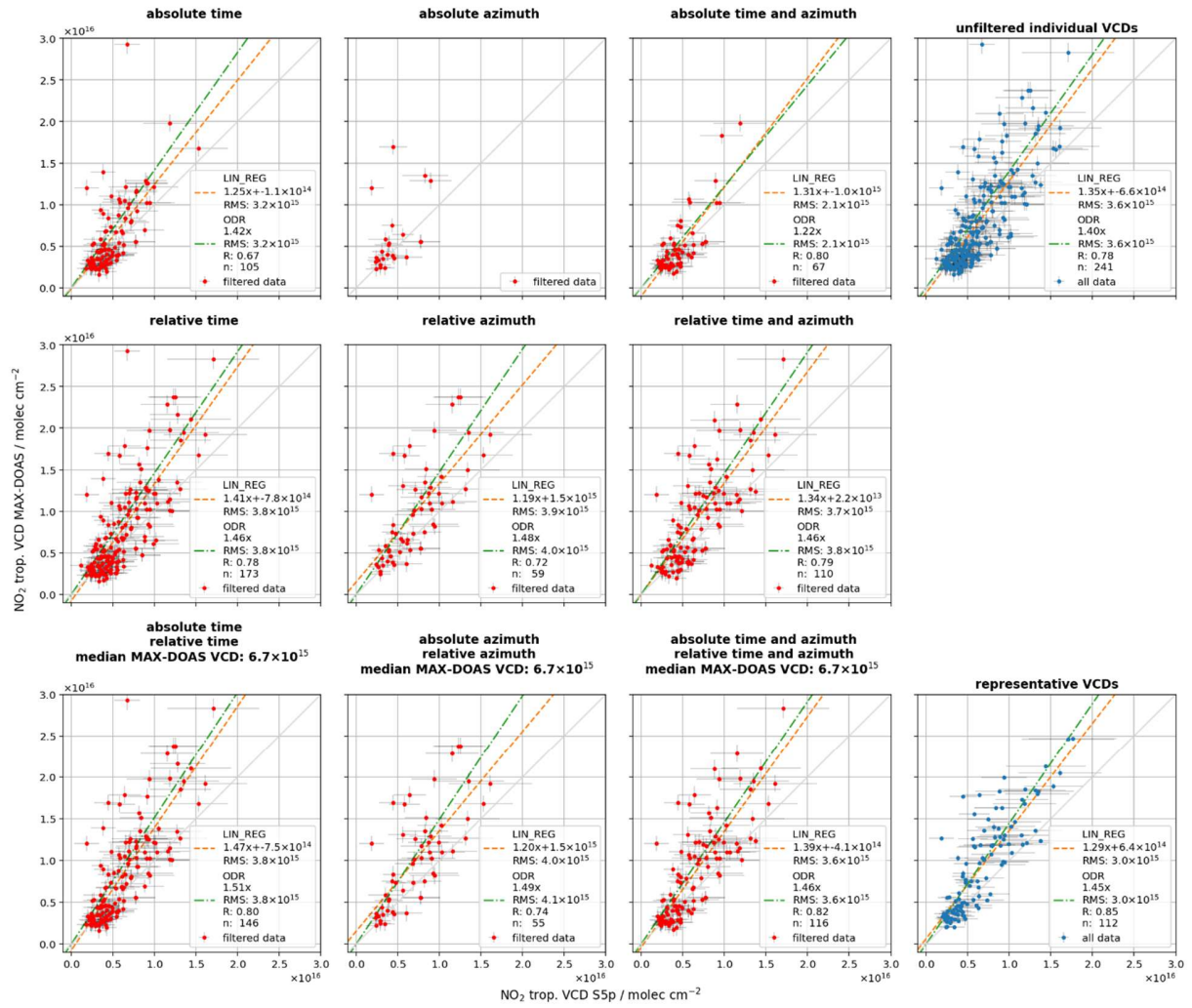


# MAINZ\_1676\_2\_pixel, MMF, max. 20 km

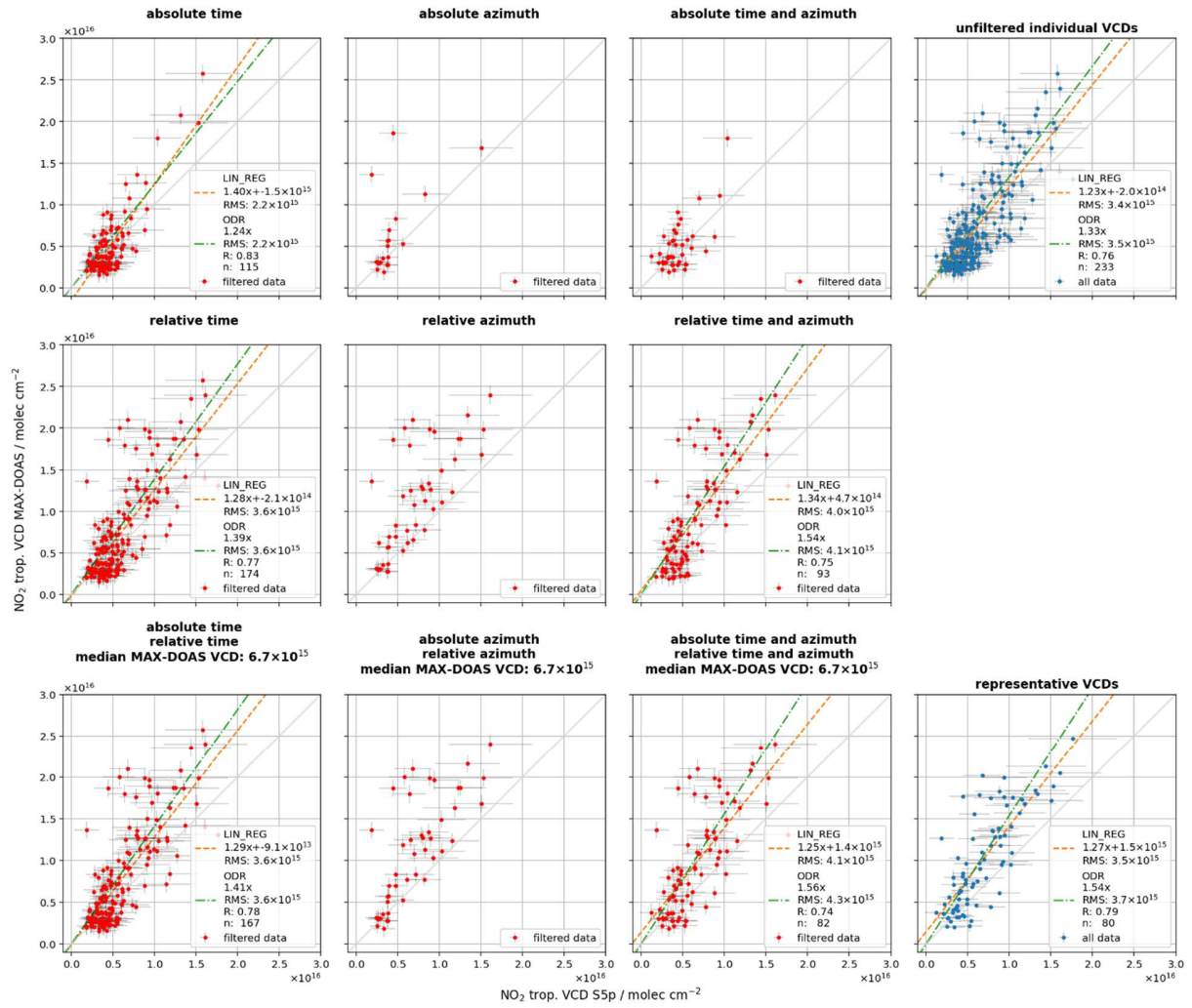




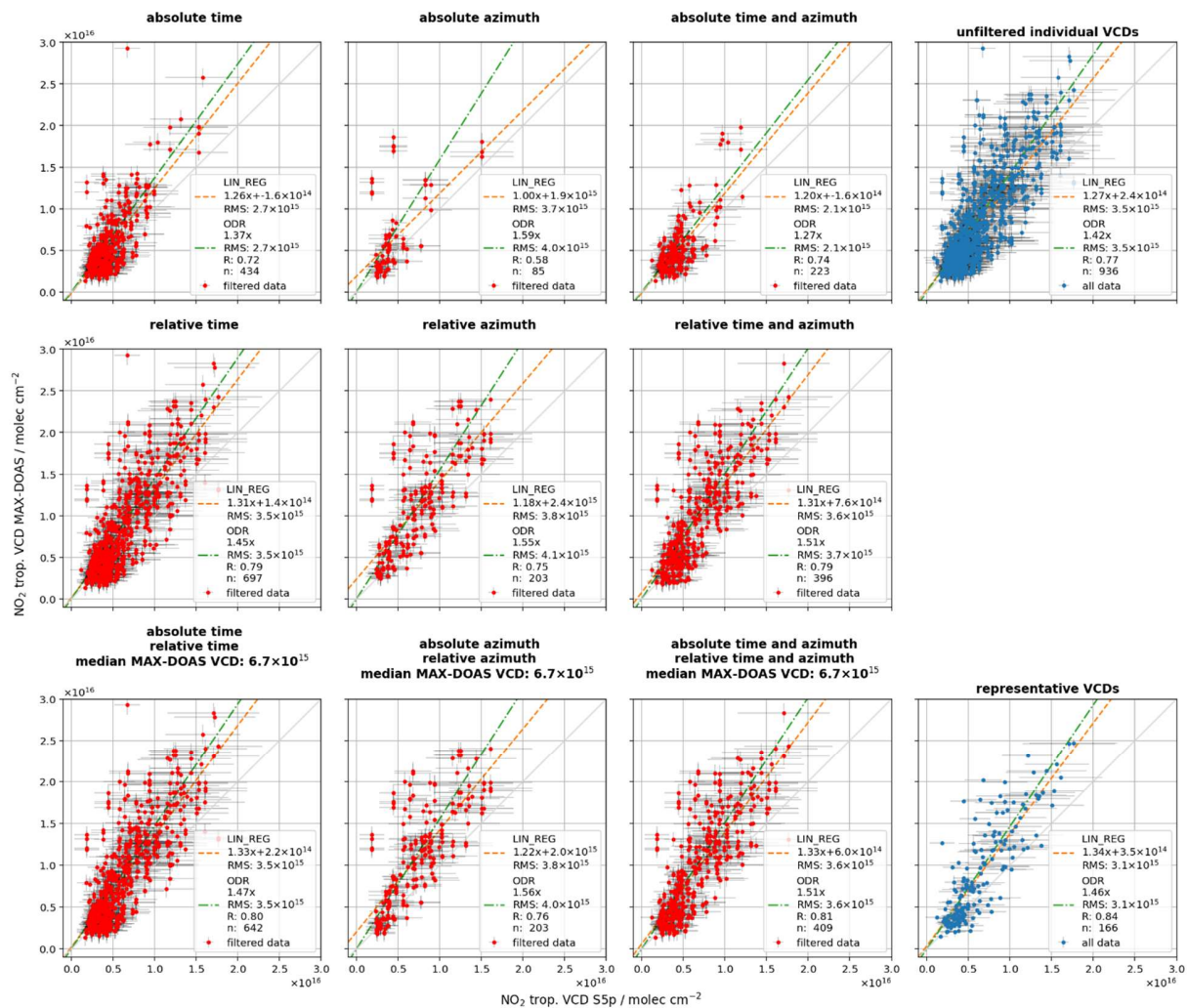
# MAINZ\_1676\_3\_pixel, MMF, max. 20 km



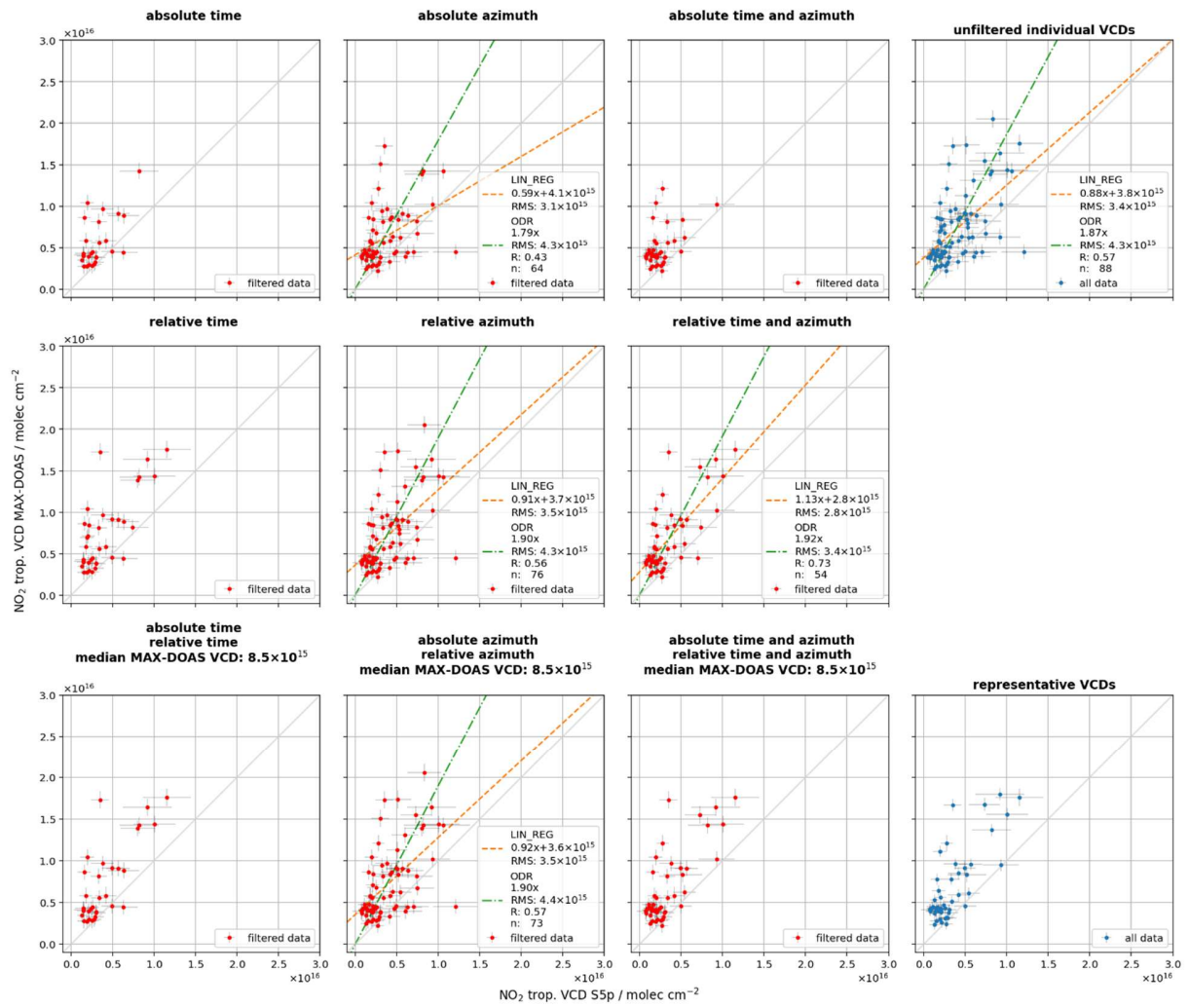
# MAINZ\_1676\_4\_pixel, MMF, max. 20 km



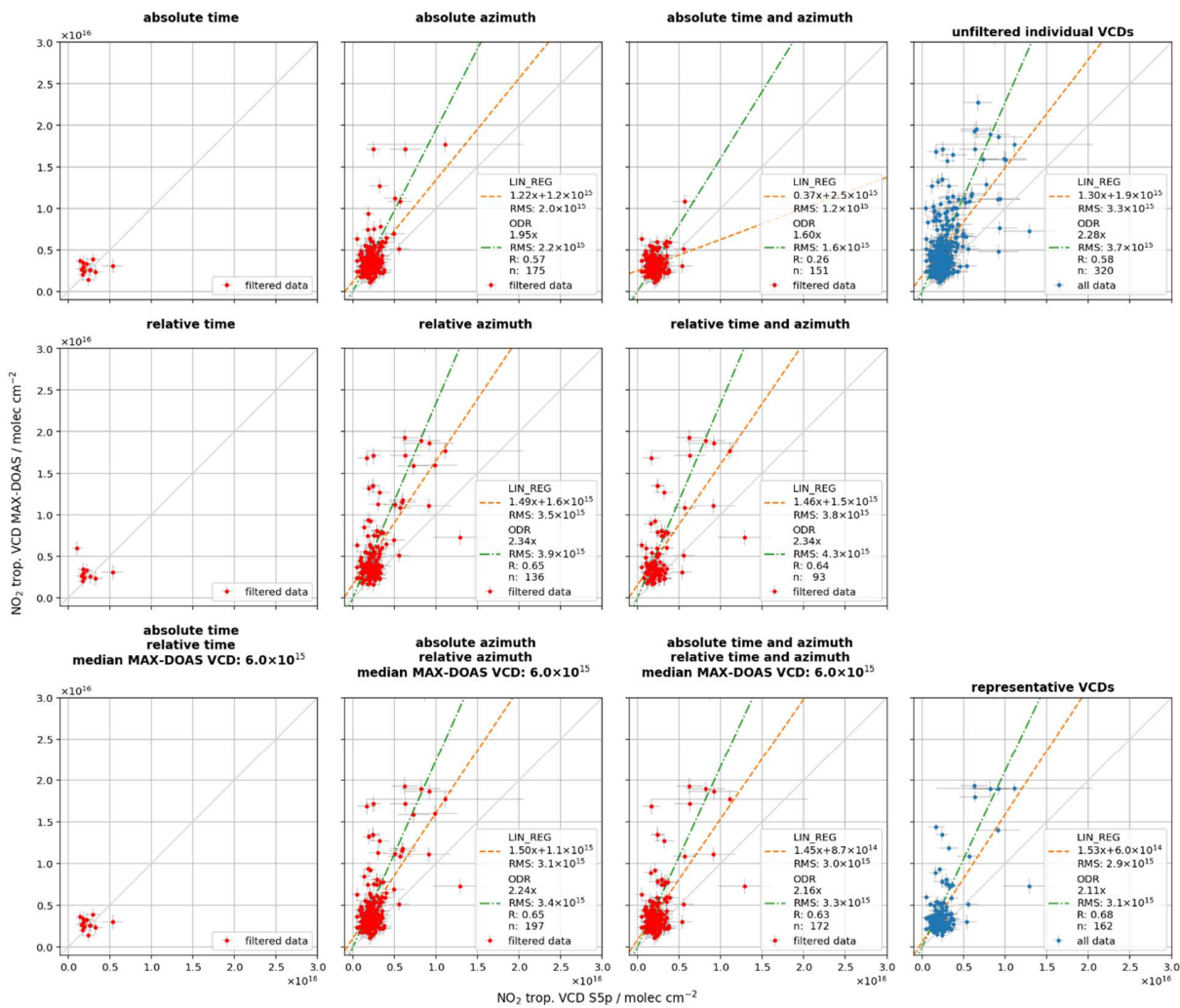
# MAINZ\_1676\_pixel, MMF, max. 20 km



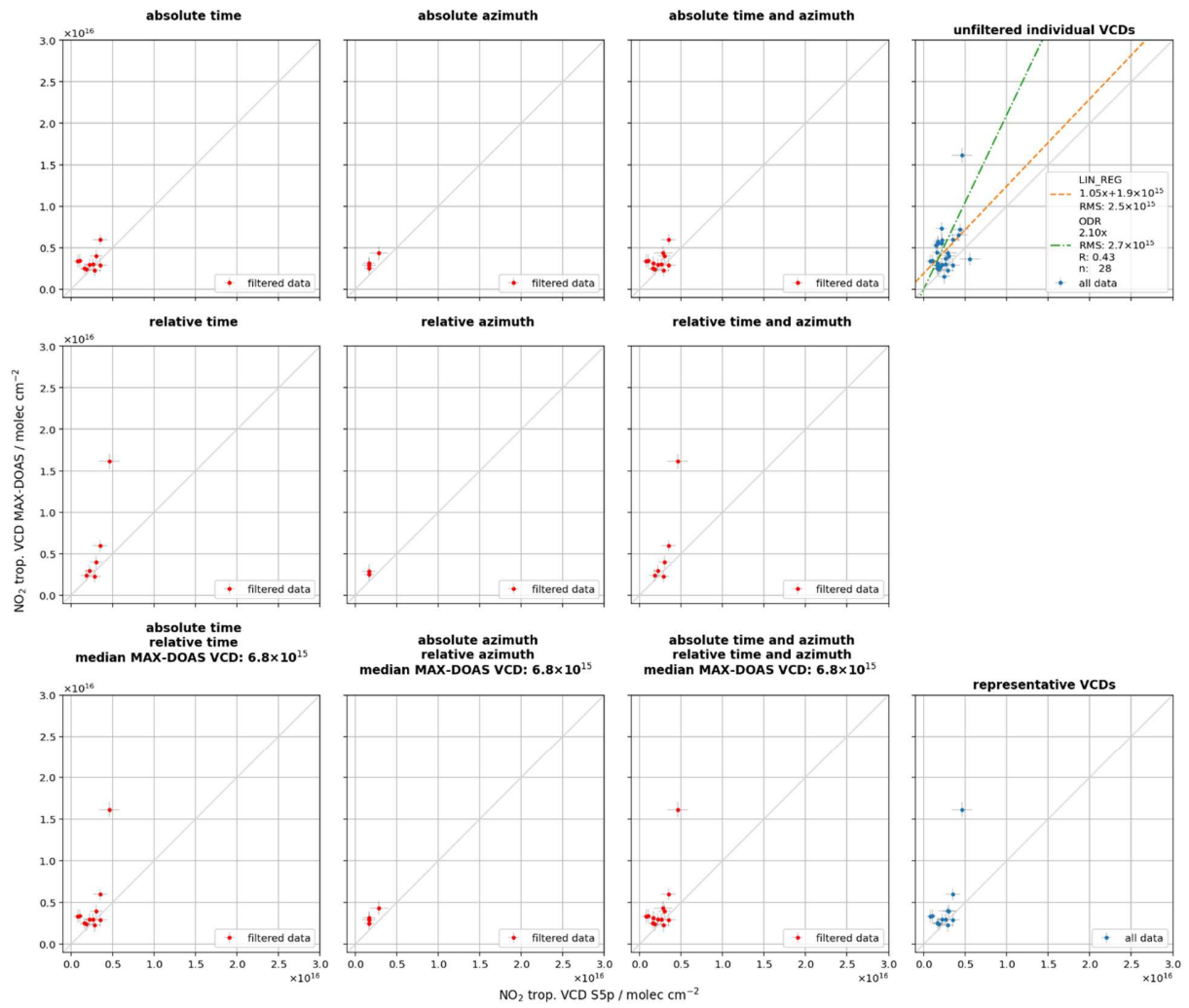
THESSALONIKI\_1683\_1\_pixel, MMF, max. 20 km



**THESSALONIKI 1686 1 pixel, MMF, max. 20 km**

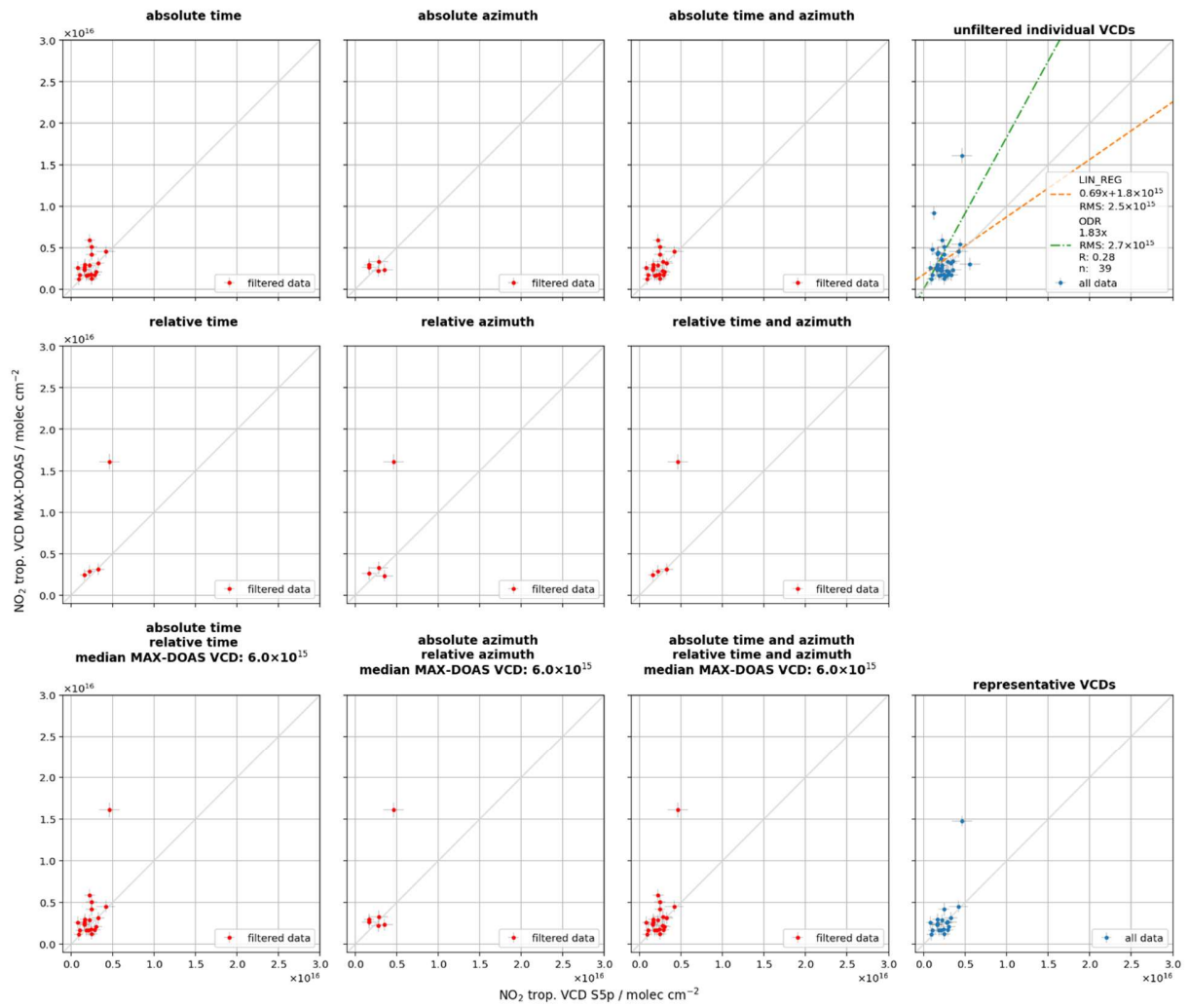


THESSALONIKI\_1699\_1\_pixel, MMF, max. 20 km



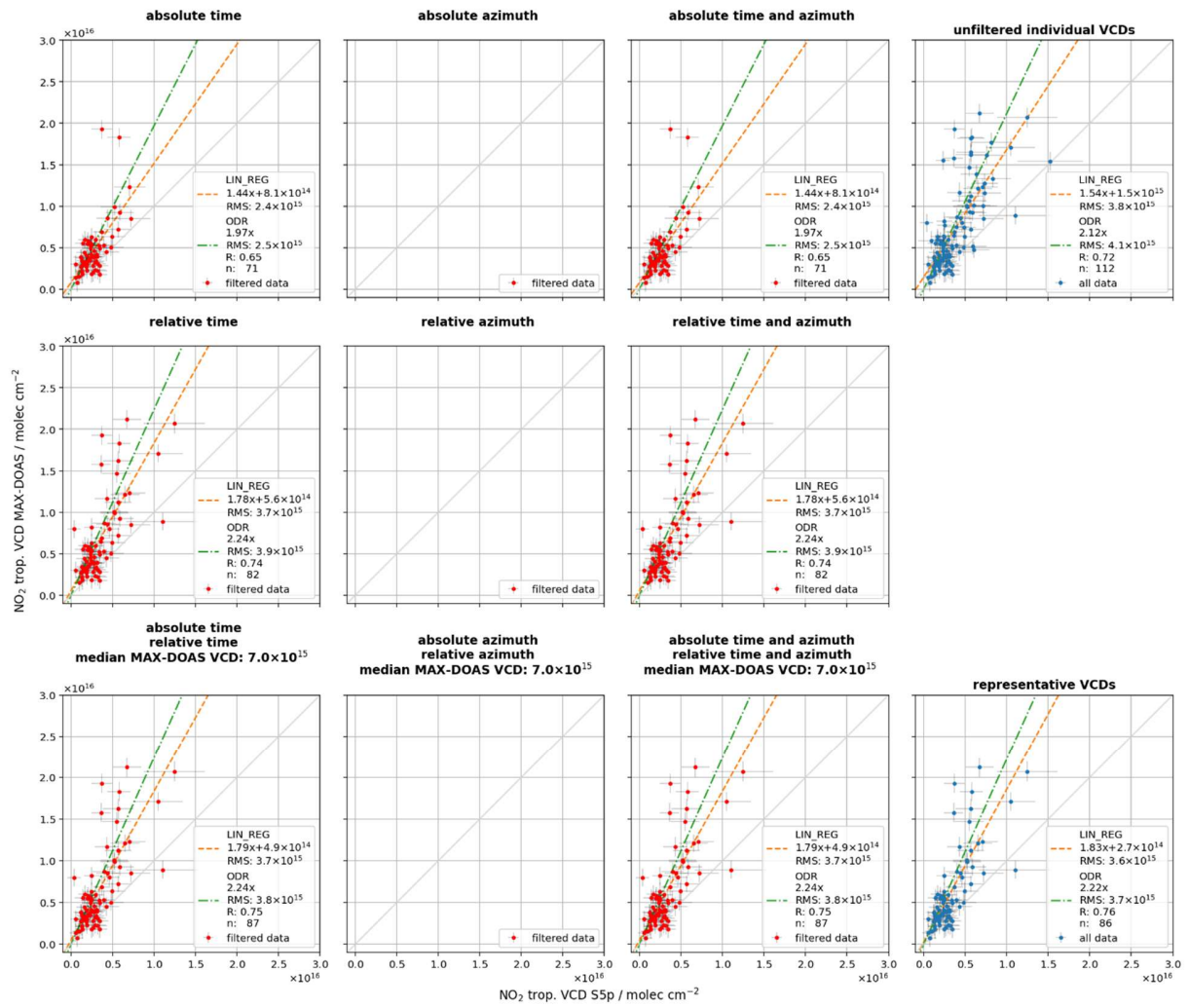


THESSALONIKI\_1699\_2\_pixel, MMF, max. 20 km

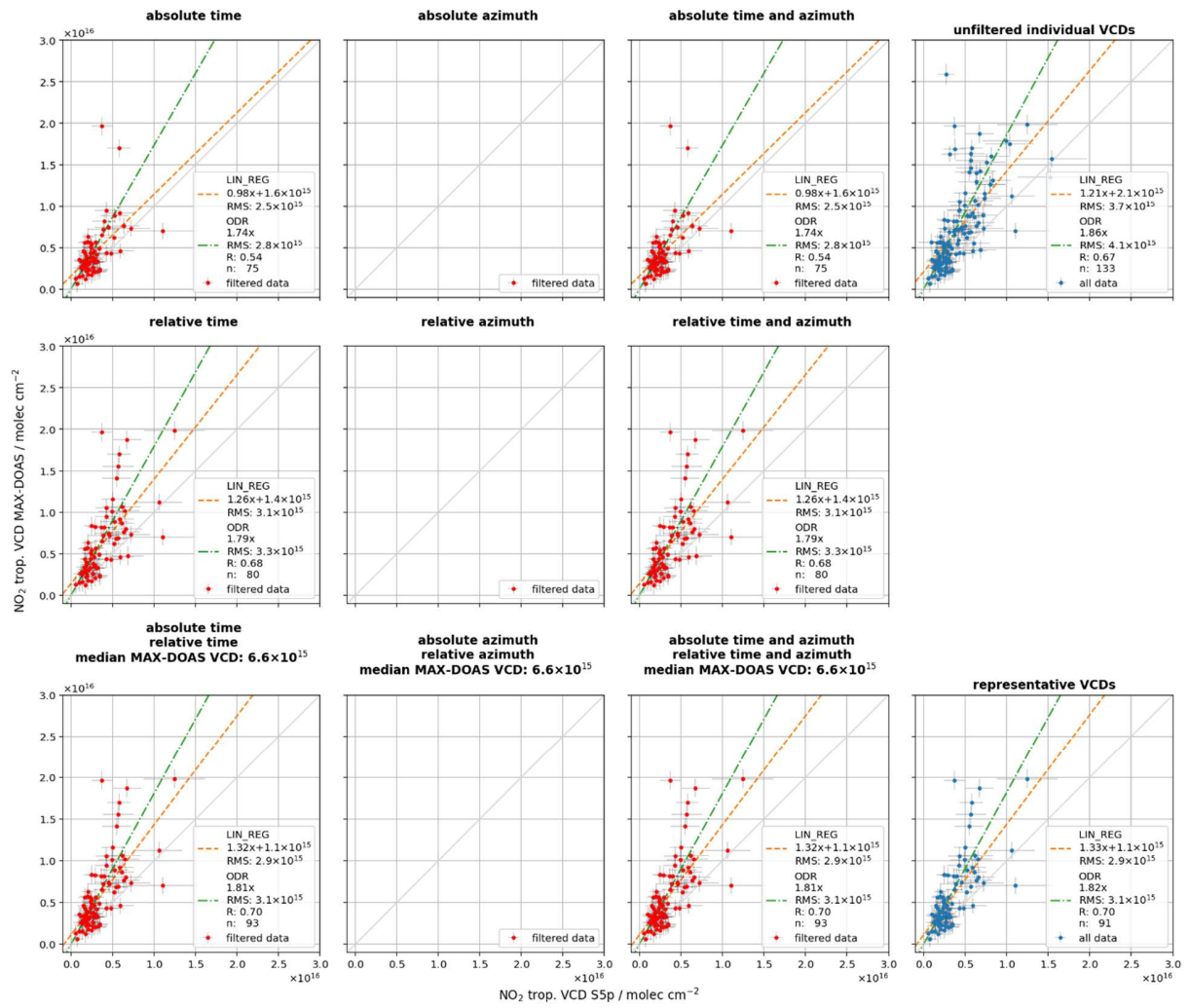




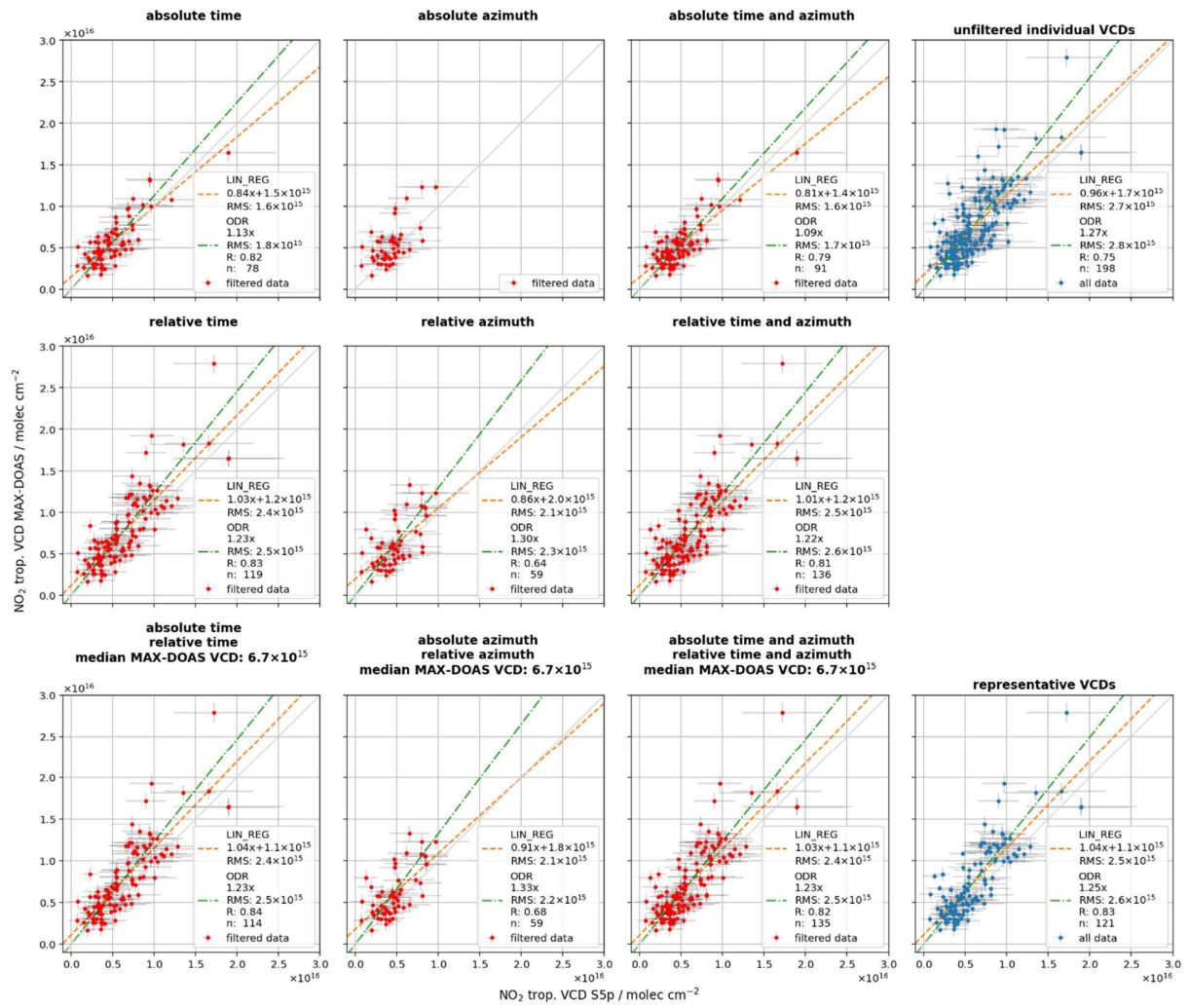
TORONTO\_1694\_1\_pixel, MMF, max. 20 km



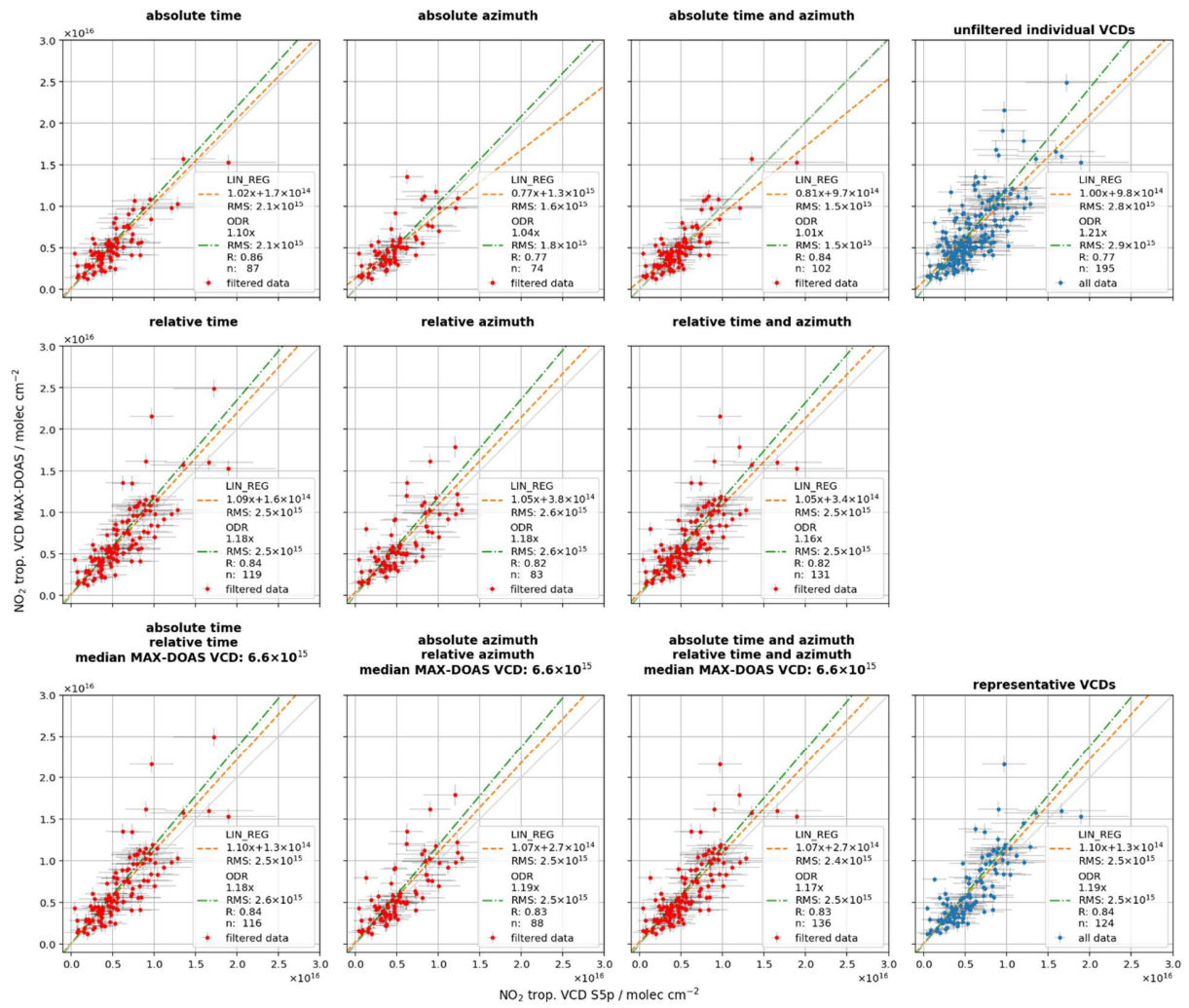
TORONTO\_1694\_2\_pixel, MMF, max. 20 km



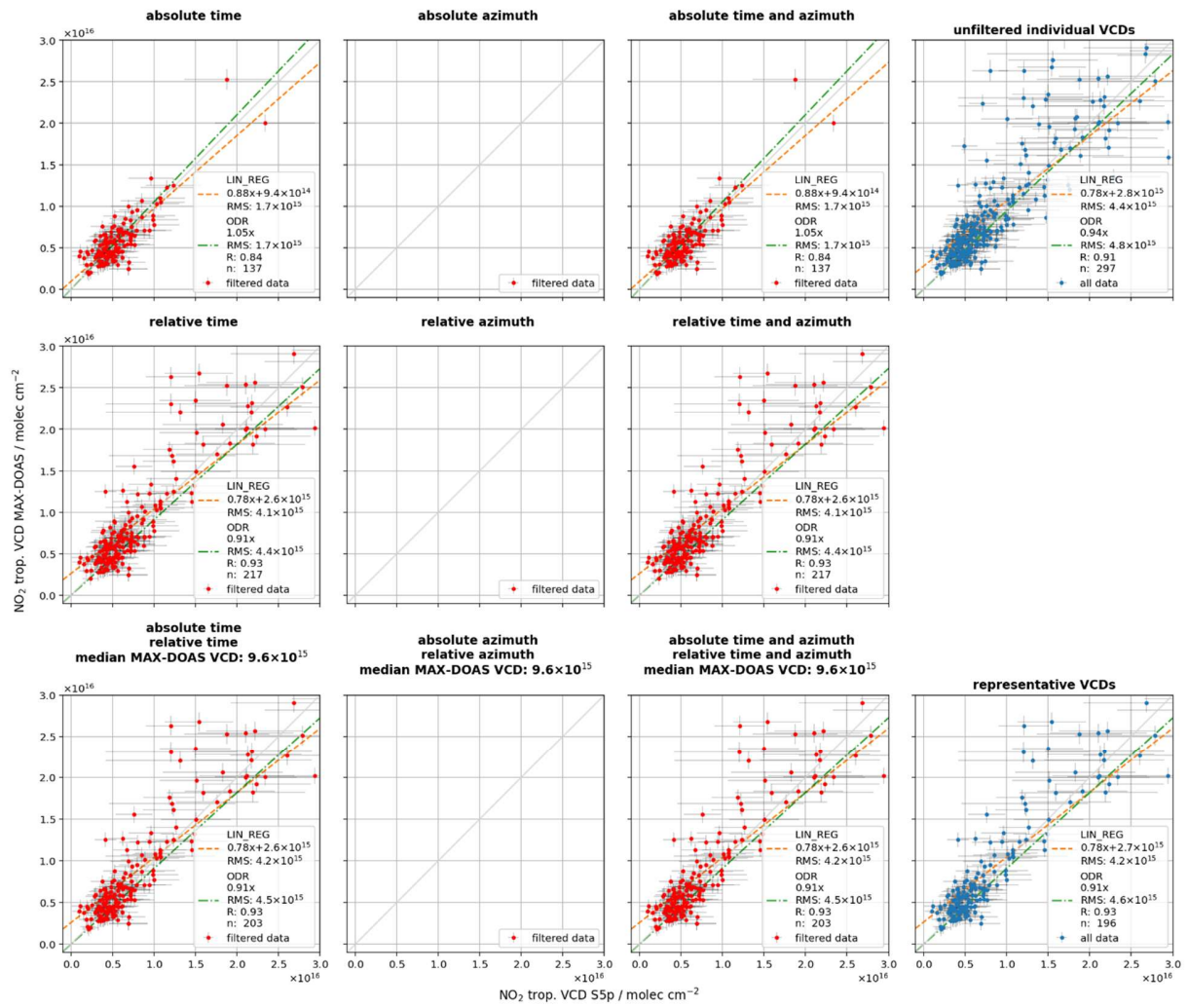
# UCCLE\_1670\_1\_pixel, MMF, max. 20 km



# UCCLE\_1670\_2\_pixel, MMF, max. 20 km



XIANGHE\_1669\_1\_pixel, MMF, max. 20 km





XIANGHE\_1669\_2\_pixel, MMF, max. 20 km

