Supplement of

Using machine learning to construct TOMCAT model and occultation measurement-based stratospheric methane (TCOM-CH4) and nitrous oxide (TCOM-N2O) profile data sets

Sandip S. Dhomse and Martyn P. Chipperfield

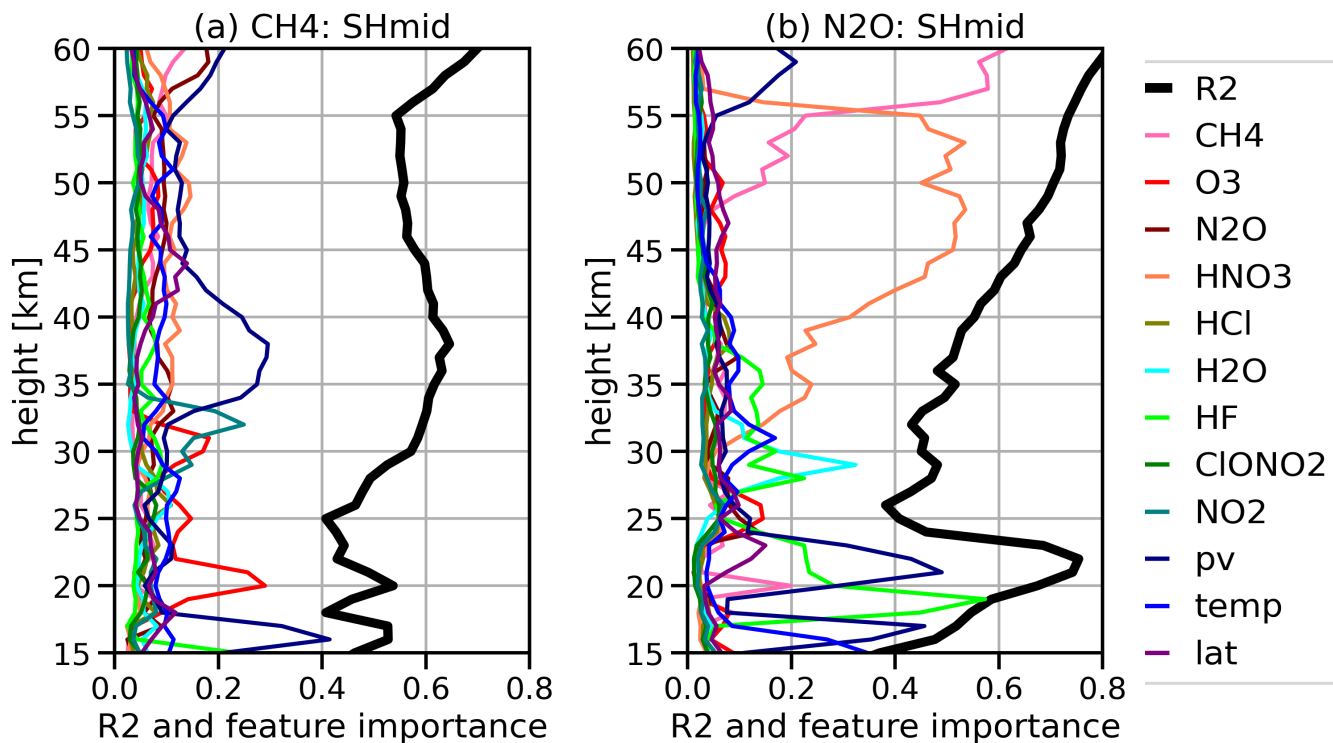*Correspondence to:* Sandip S. Dhomse (s.s.dhomse@leeds.ac.uk)

Figure S1. Vertical profiles of the variance (R^2) and feature importances estimated by XGBoost regression models for the TOMCAT-observation differences for (a) CH_4 (1991-2018) and (b) N_2O (2004-2018, ACE only) for the South Hemisphere mid-latitude (SHmid, $20°S$–$70°S$) latitude bin. See equation 1 and subsequent information in the manuscript about the features (total 13) or variables used in the XGBoost regression model.
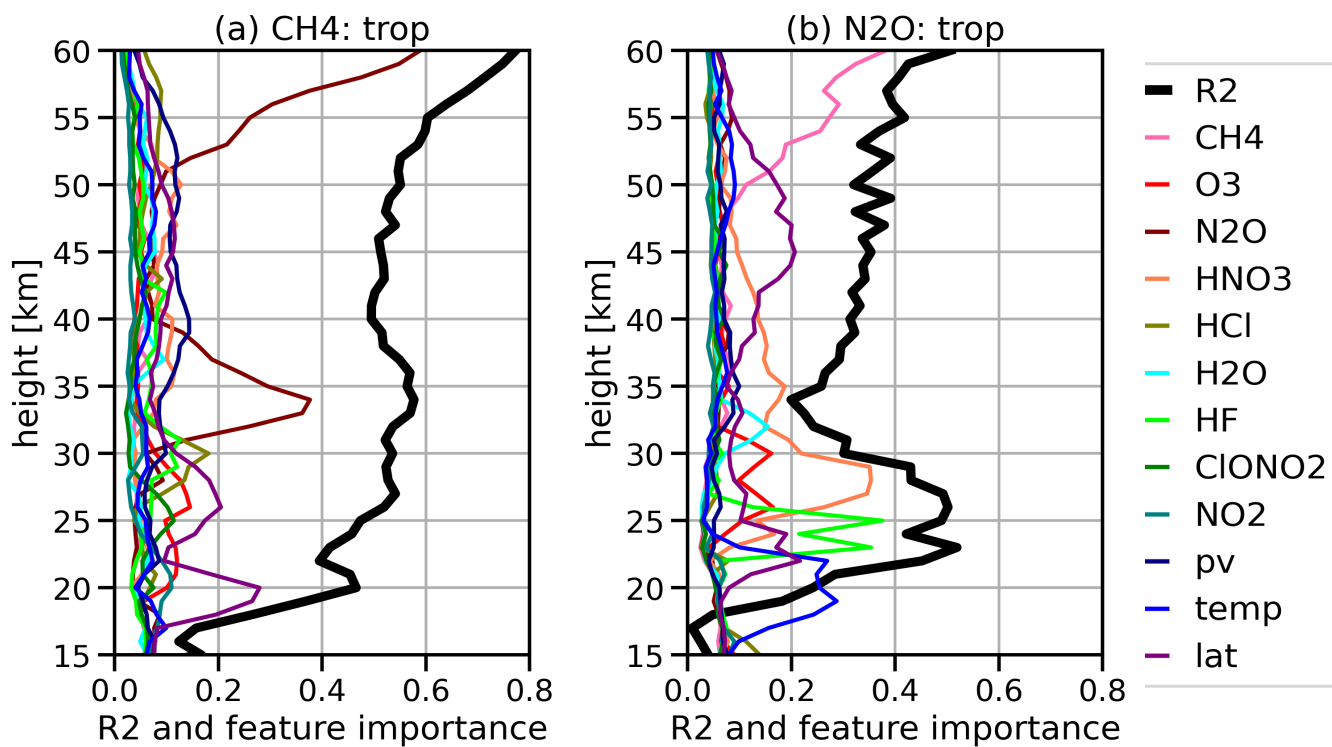
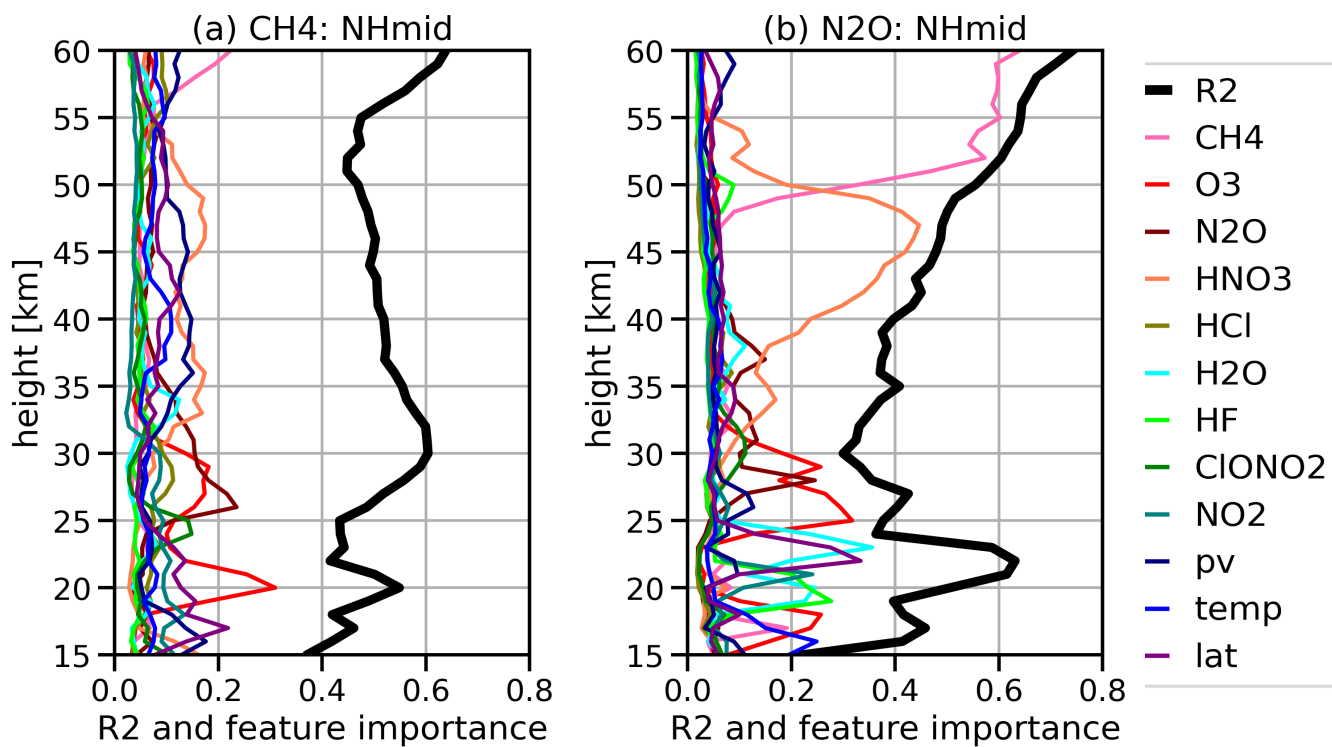**Figure S2.** Same as S1, but for tropical latitude band (40° S–40° N)

.

Figure S3. Same as S1, but for northern hemisphere mid-latitude (NHmid) band (20°N–70°N)
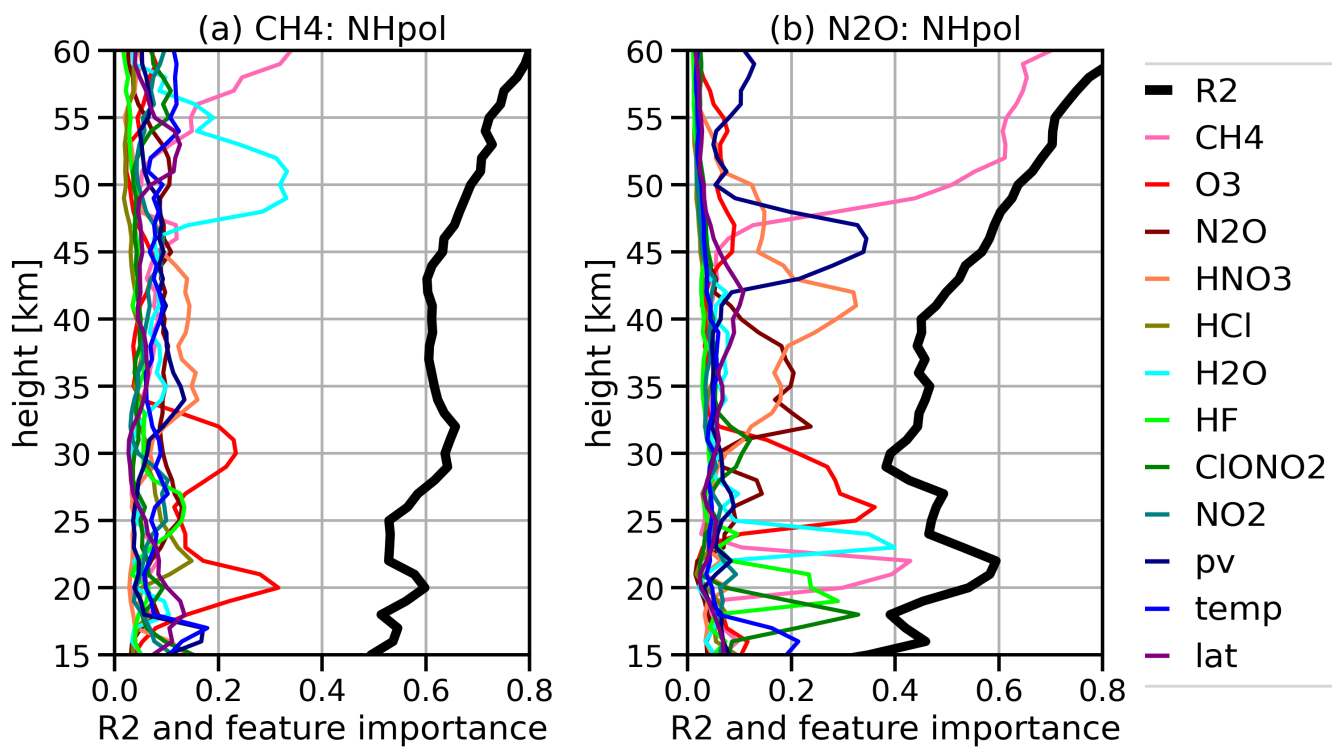
.

Figure S4. Same as S1, but for northern hemisphere polar (NHPol) latitude band (50°N–90°N)
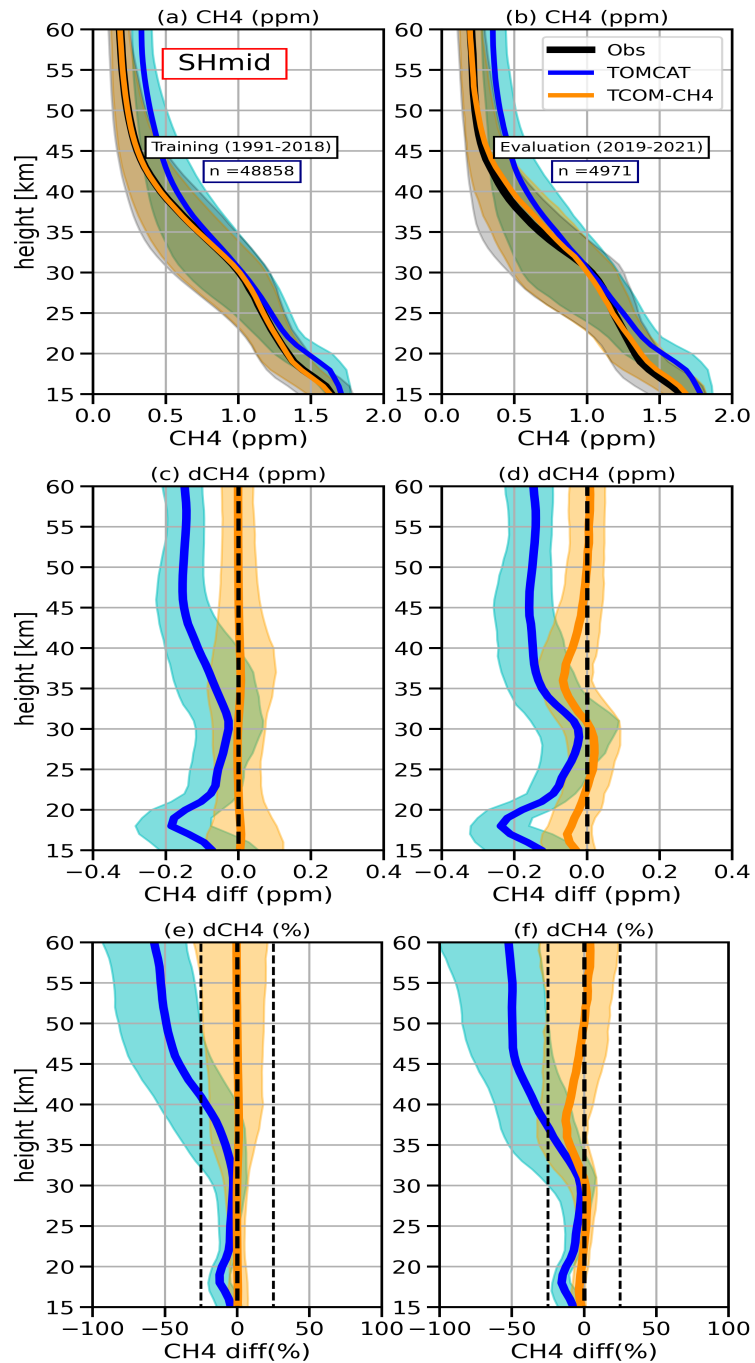
.

Figure S5. Panels (a) and (b). Comparison between TOMCAT (blue), TCOM-CH4 (orange) and satellite measurement-based (black) CH_4 profiles for SHmid (20°S–70°S) latitude band. Solid lines indicate median profiles while shaded regions show 10th and 90th percentile range. Comparisons are shown for training (1992-2018) and evaluation (2019-2021) periods in panels (a, left) and (b, right), respectively. Panels (c) - (f). Differences between TOMCAT and TCOM-CH4 w. r. t. satellite data sets in absolute units (ppm) and percent. Right (c and e) and left (d and f) panels show differences for the training (1992-2018) and evaluation (2019-2021) periods.
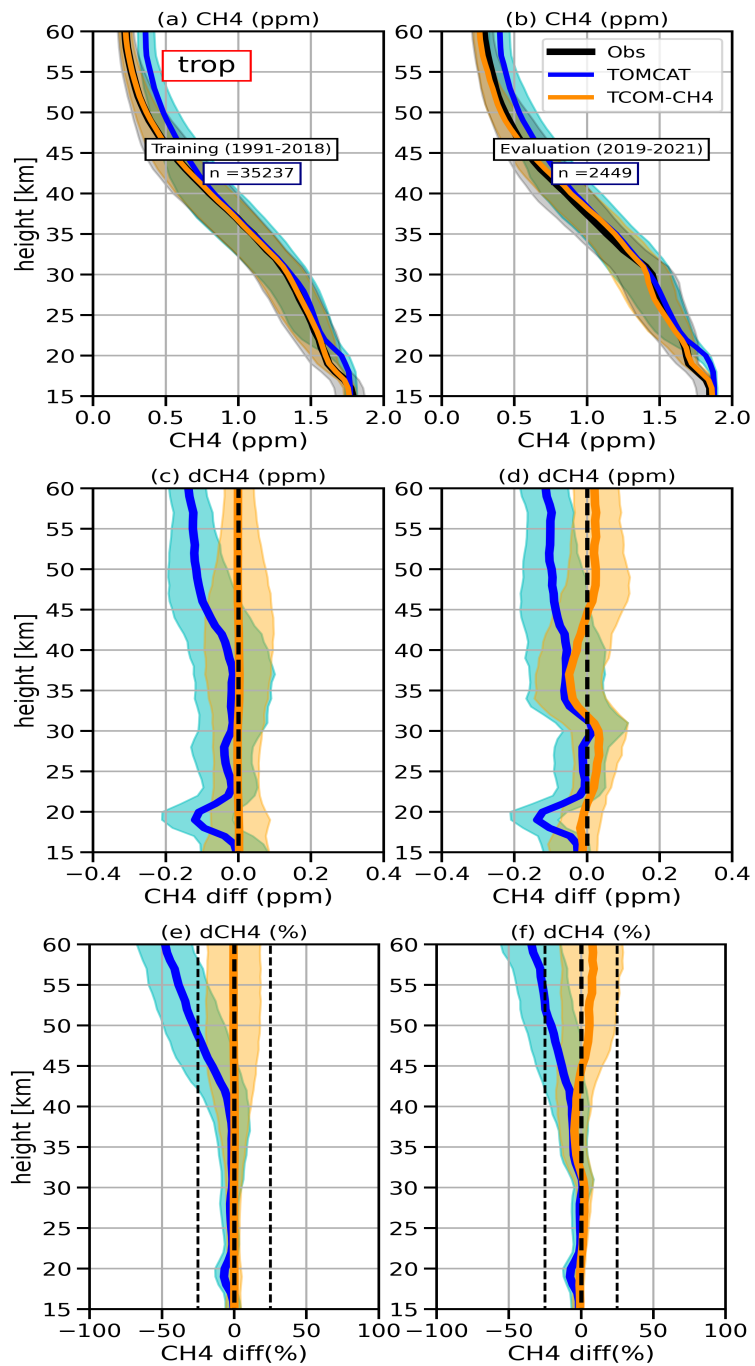
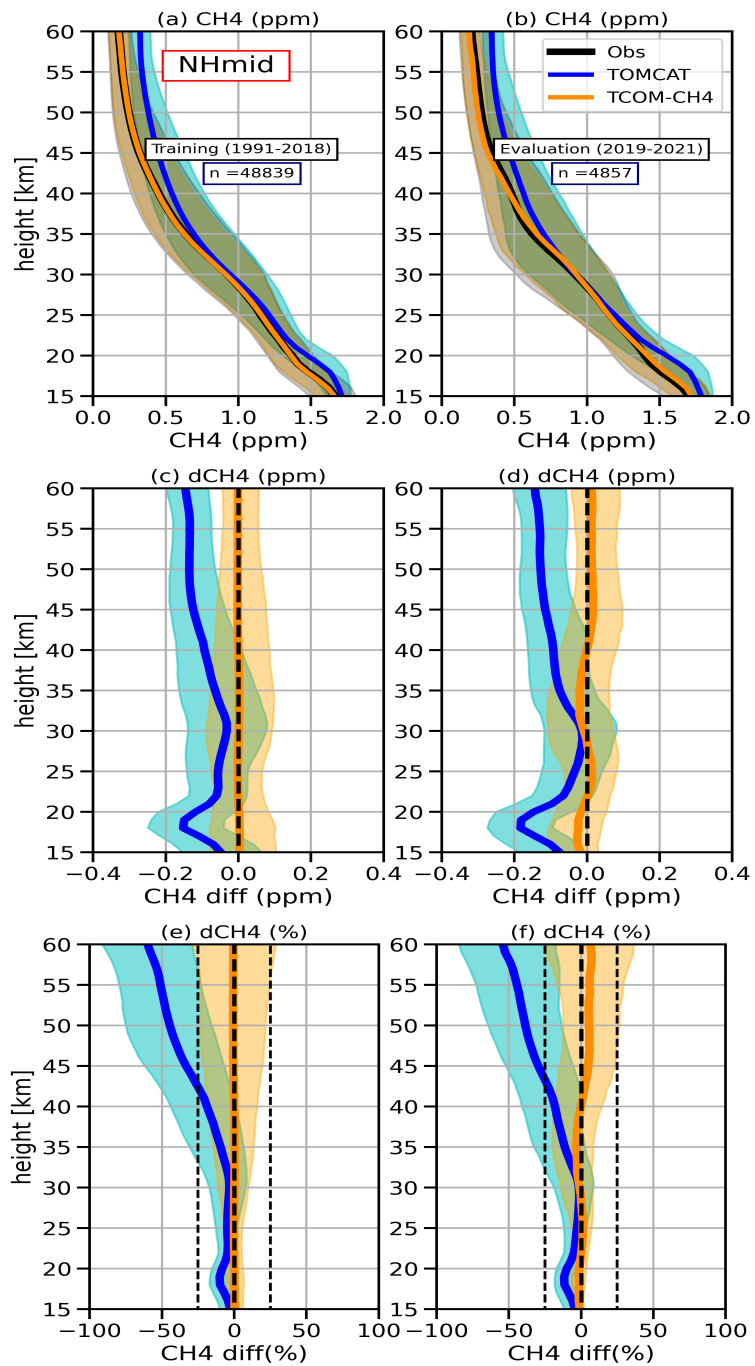**Figure S6.** Same as S5, but for tropical (trop) latitude band (40°S–40°N)

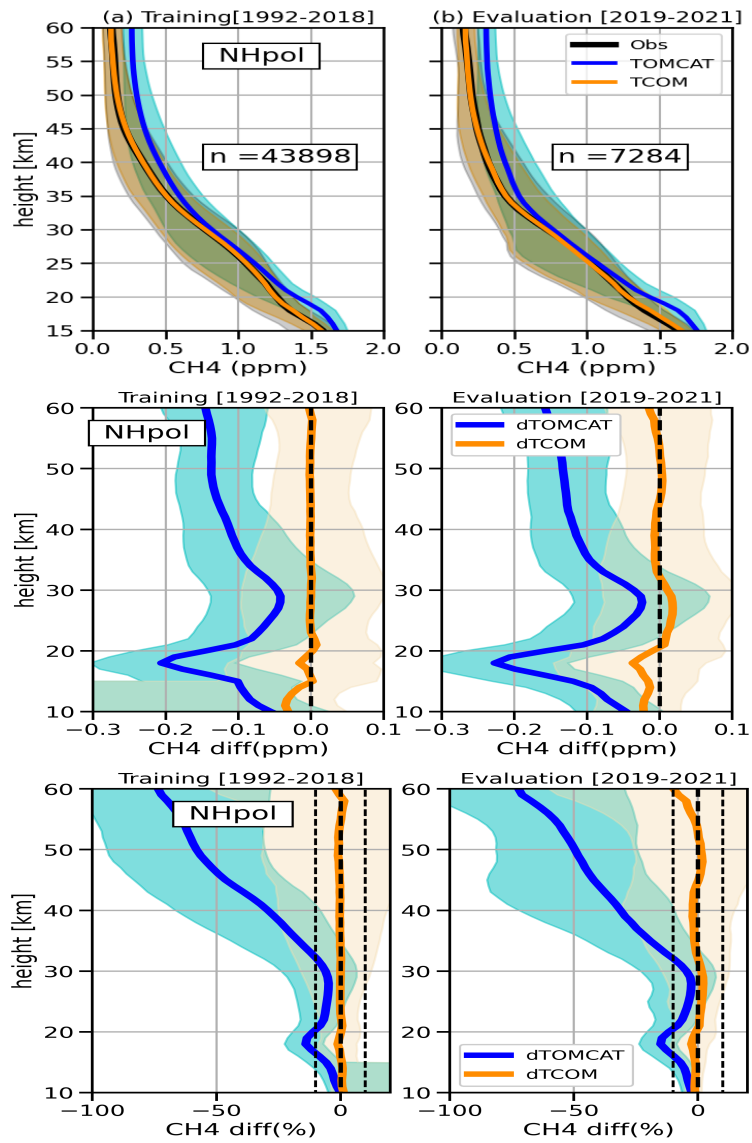**Figure S7.** Same as S5, but for northern hemisphere mid-latitude (NHmid) band (20°N–70°N)

.

Figure S8. Same as S5, but for northern hemisphere polar-latitude (NHpol) band (50°N–90°N)
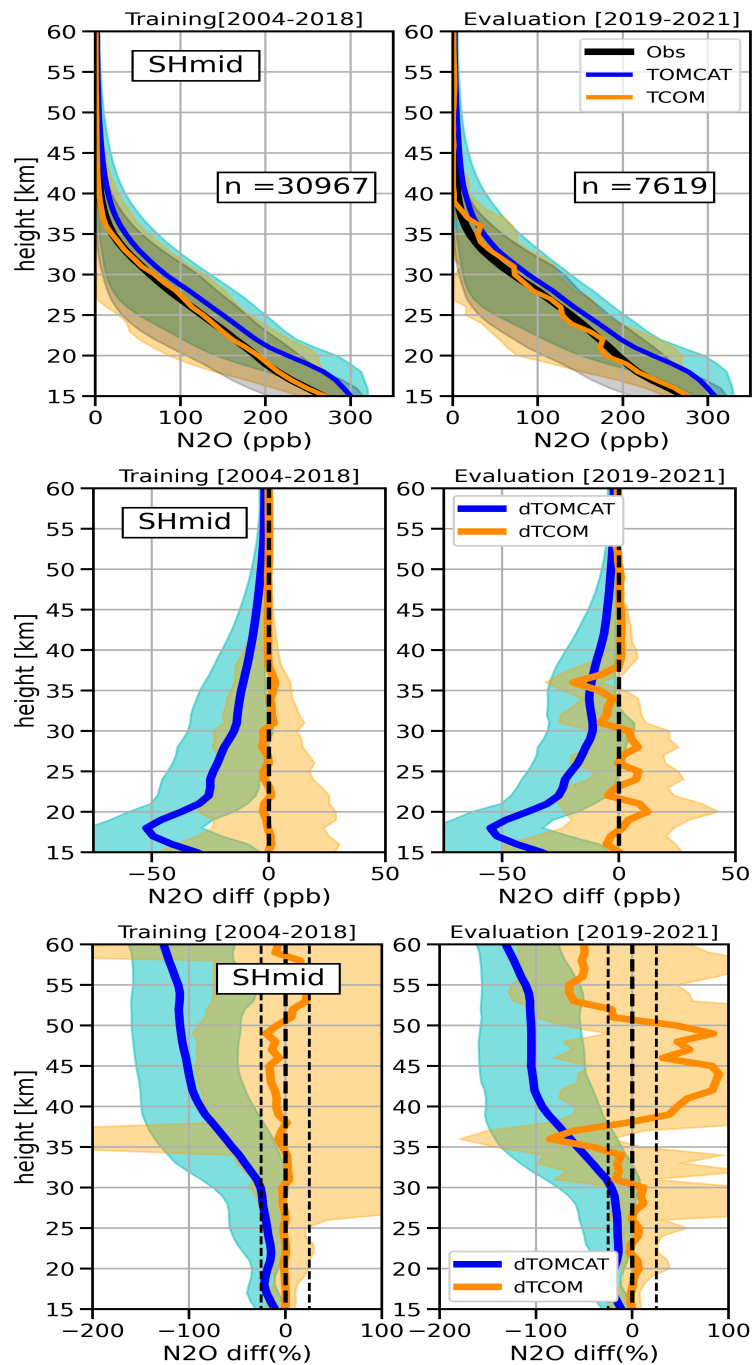
.

'

Figure S9. Panels (a) and (b). Comparison between TOMCAT (blue), TCOM-N2O (orange) and ACE-FTS satellite (black) N_2O profiles measurements for SHmid (20°S–70°S) latitude band. Solid lines indicate median profiles while shaded regions show 10th and 90th percentile range. Comparisons are shown for training (1992-2018) and evaluation (2019-2021) periods in panels (a, left) and (b, right), respectively. Panels (c) - (f). Differences between TOMCAT and TCOM-N2O w. r. t. satellite data sets in absolute units (ppm) and percent. Right (c and e) and left (d and f) panels show differences for the training (1992-2018) and evaluation (2019-2021) periods.
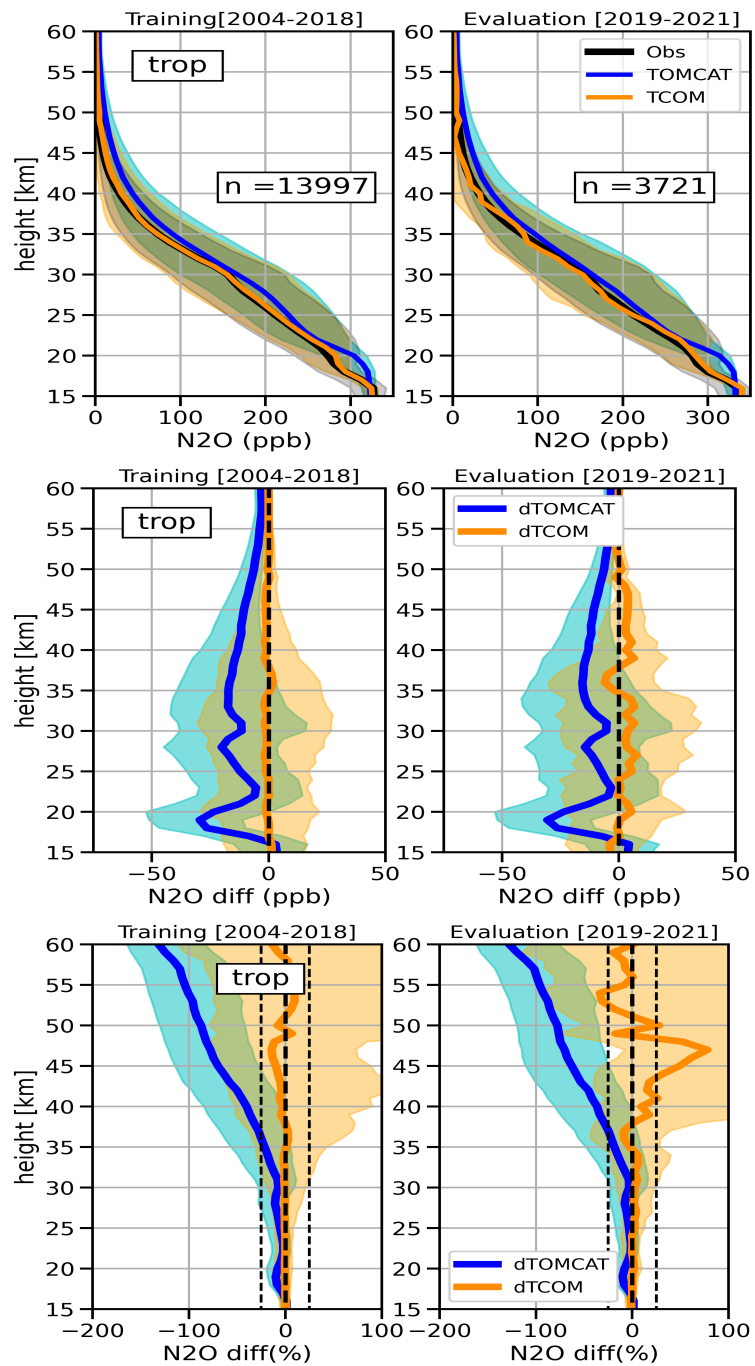
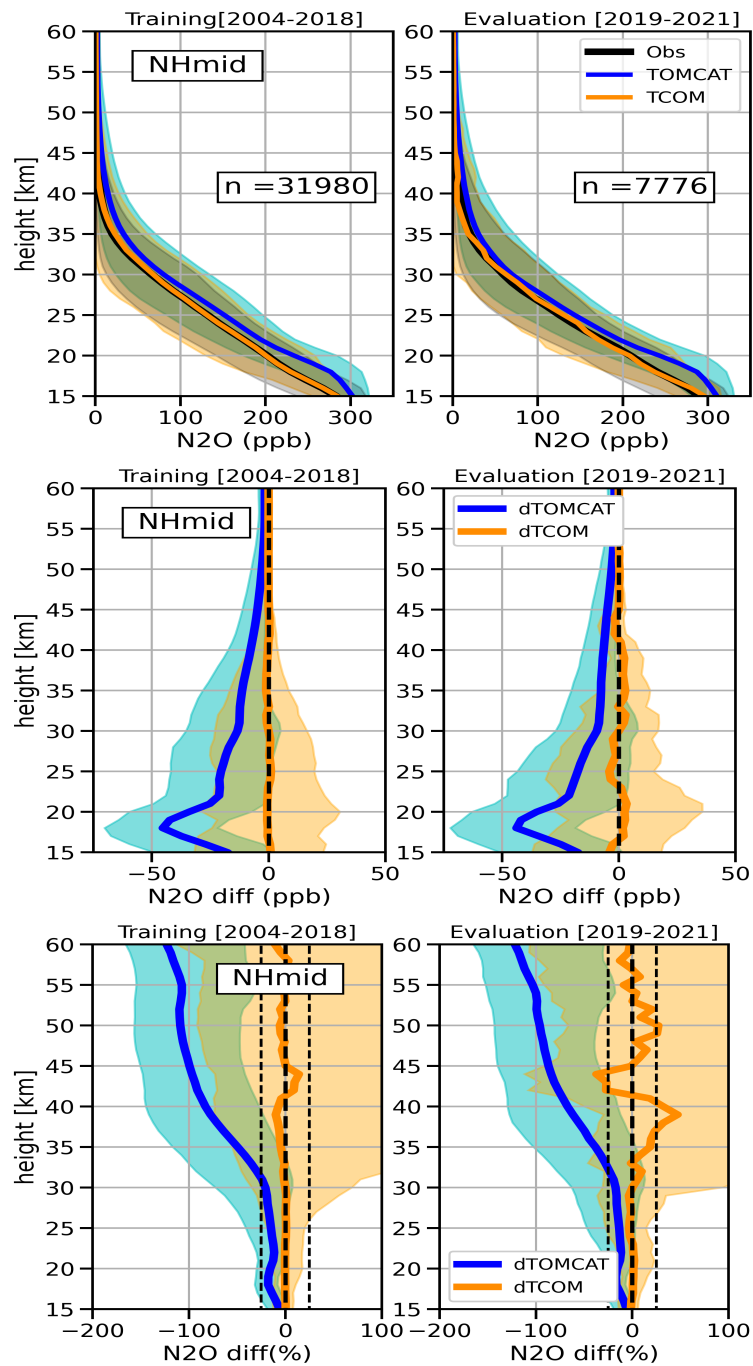**Figure S10.** Same as S9, but for tropical (trop) latitude band (40°S–40°N)

.

11

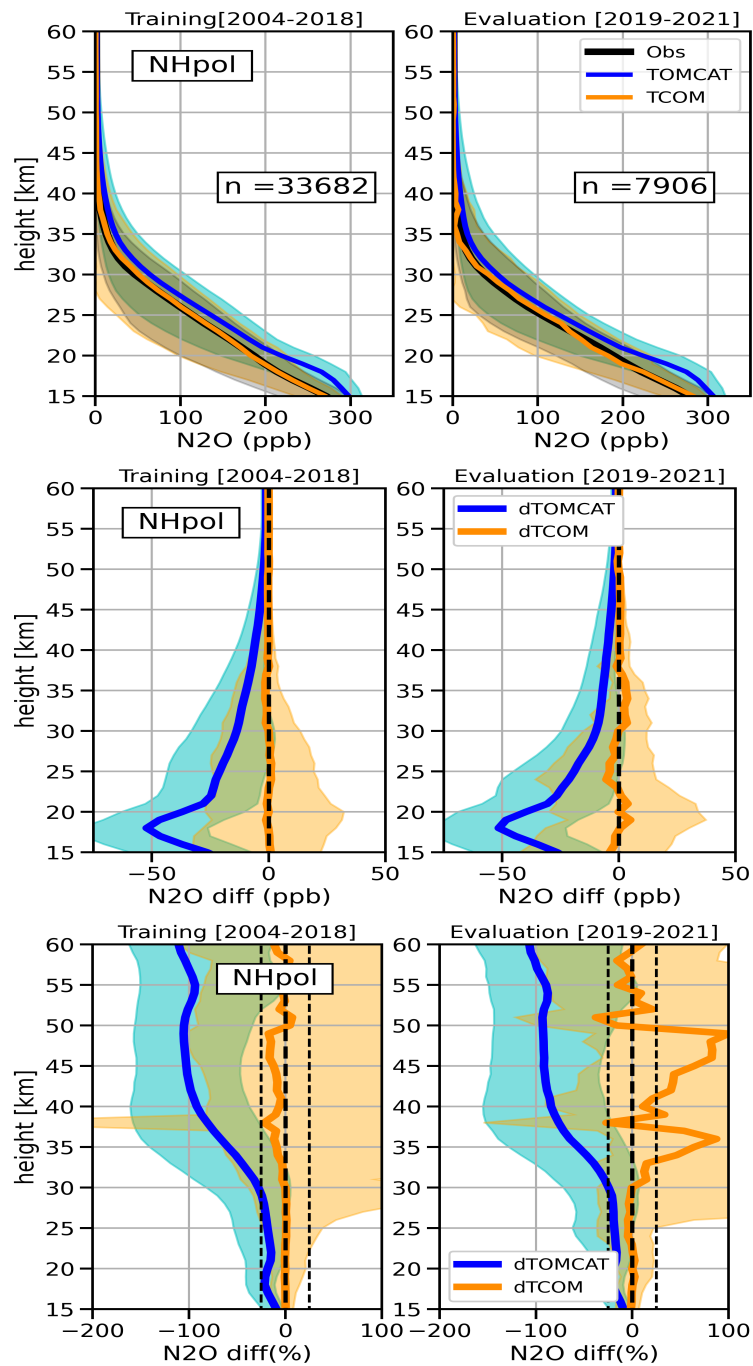**Figure S11.** Same as S9, but for northern hemisphere mid-latitude (NHmid) band (20°N–70°N)

.

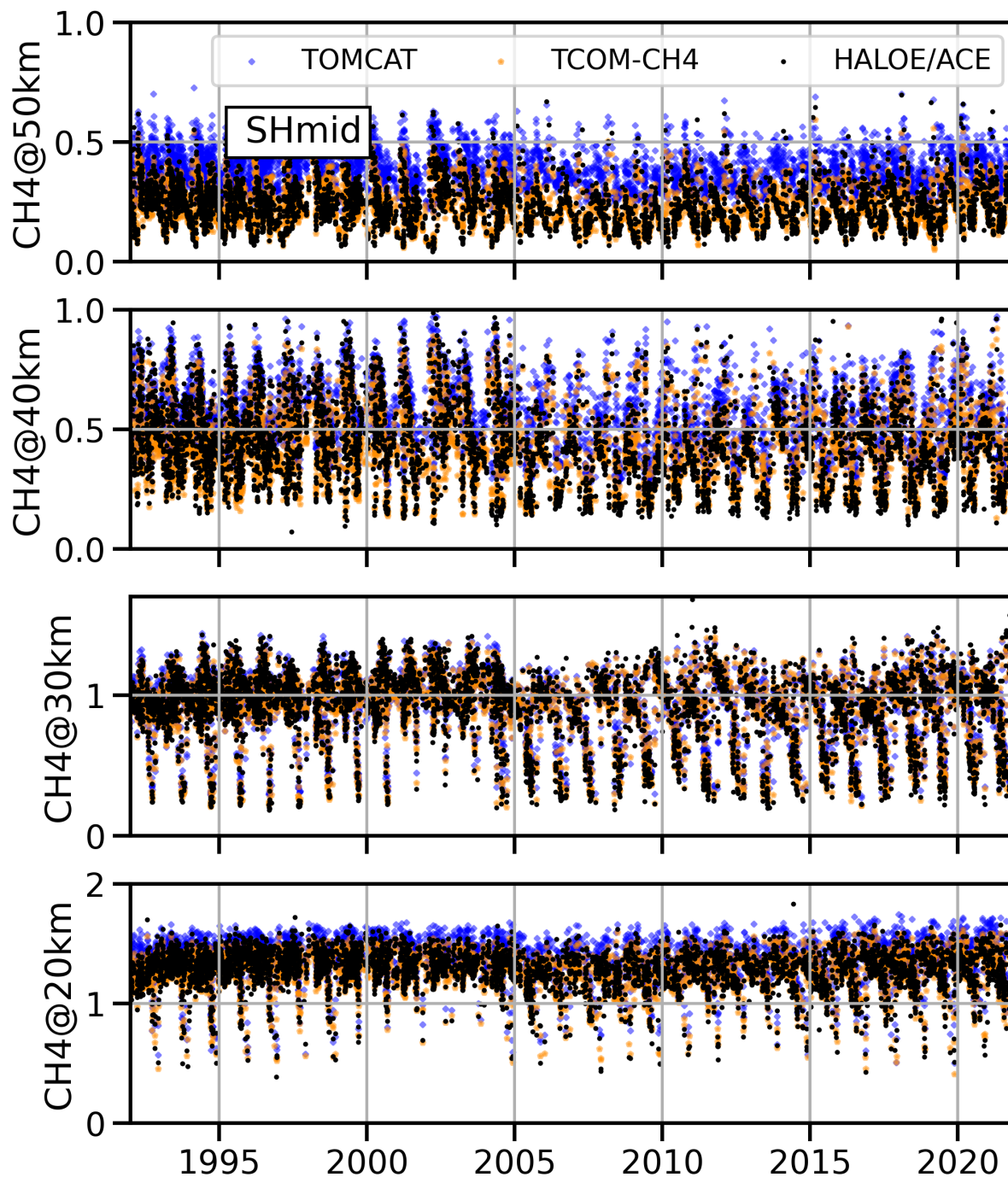**Figure S12.** Same as S9, but for northern hemisphere polar-latitude (NHpol) band (50°N–90°N)

.

Figure S13. Time evolution (1992-2021) of CH_4 from TOMCAT (blue crosses), TCOM-CH4 (orange diamonds) and satellite data (black dots) for SHmid ($20°S–70°S$) at 20, 30, 40 and 50 km. Note that for clarity only 10% (every 10th) of data points are shown and due to sharp gradient in vertical distribution, the y axis range varies between the panels.
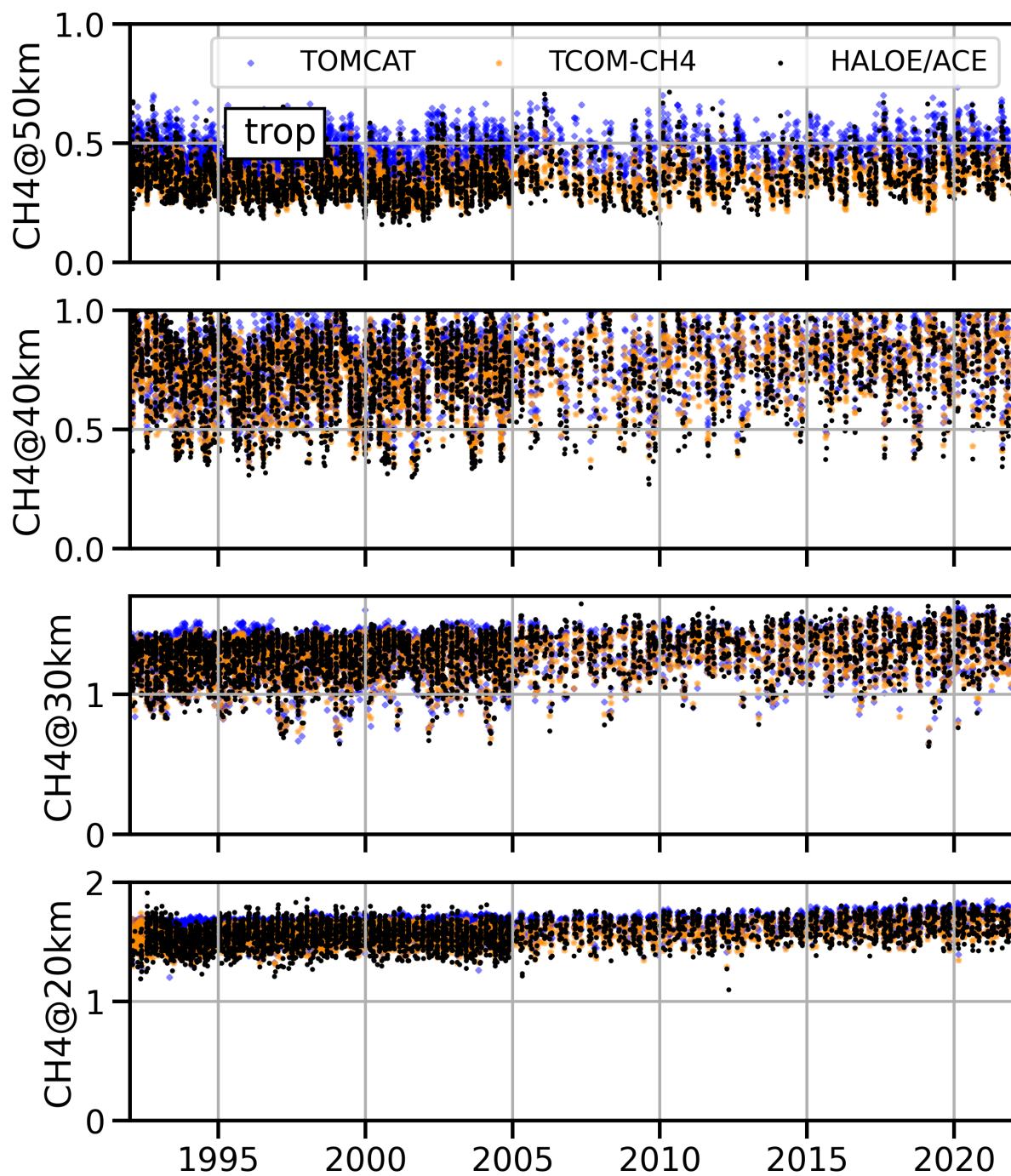
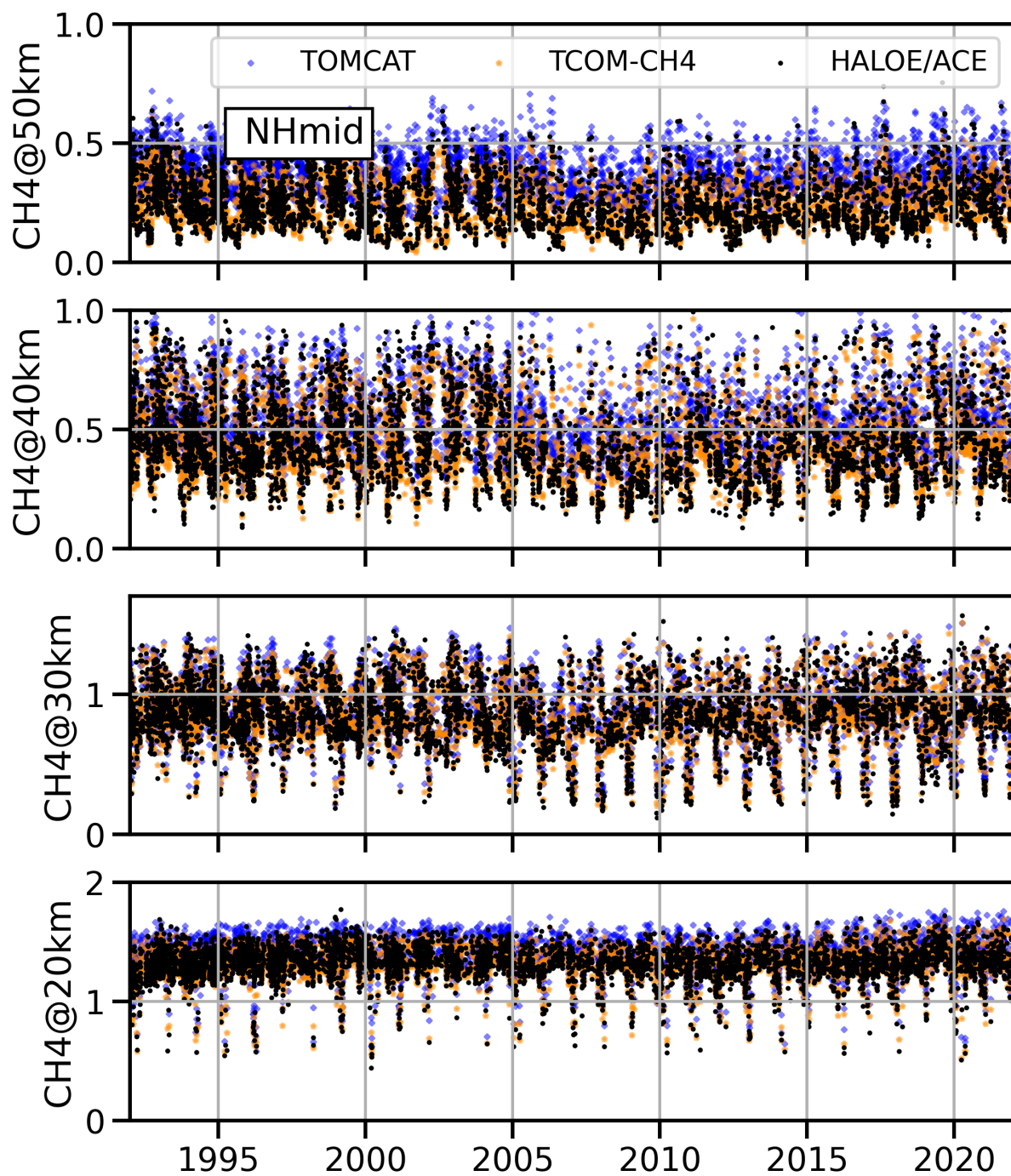**Figure S14.** Same as S13, but for tropical (trop) latitude band (40°S–40°N)

.

Figure S15. Same as S13, but for northern hemisphere mid-latitude (NHmid) band (20°N–70°N)
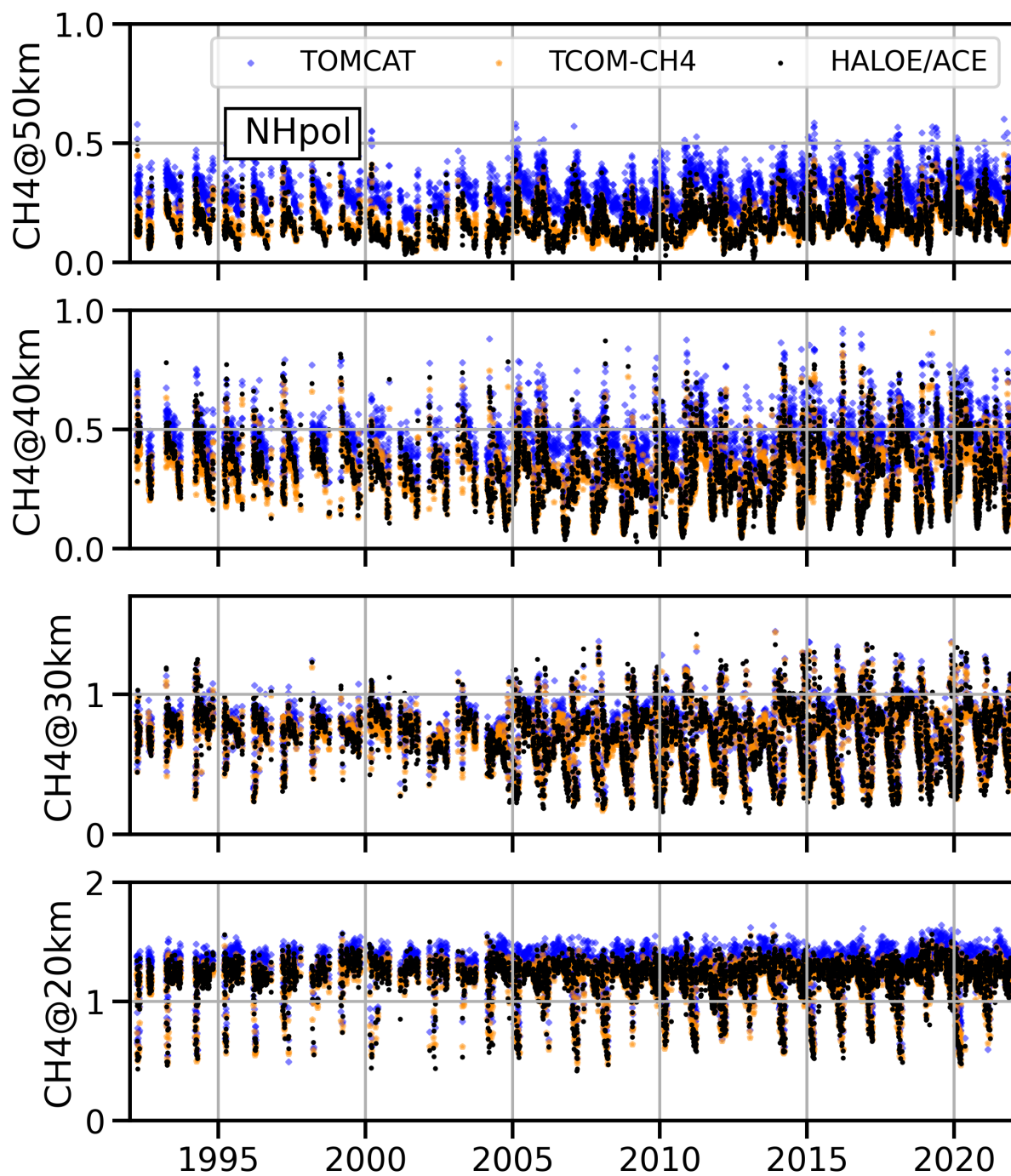
.

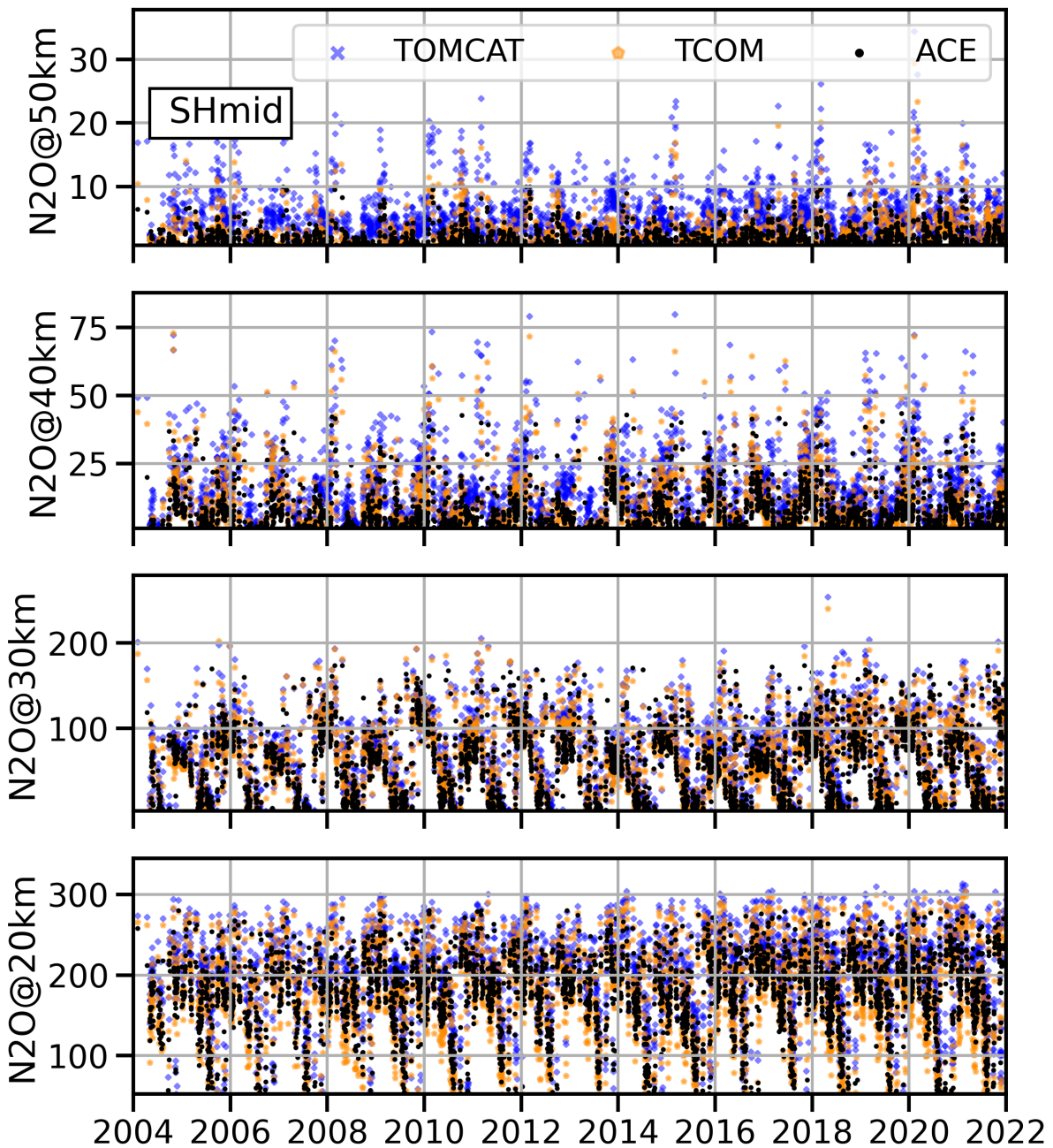**Figure S16.** Same as S13, but for northern hemisphere polar-latitude (NHpol) band (50°N–90°N)

.

Figure S17. Time evolution (1992-2021) of CH_4 from TOMCAT (blue crosses), TCOM-CH4 (orange diamonds) and satellite data (black dots) for SHmid ($20°S$–$70°S$) at 20, 30, 40 and 50 km. Note that for clarity only 10% (every 10th) of data points are shown and due to sharp gradient in vertical distribution, the y axis range varies between the panels.
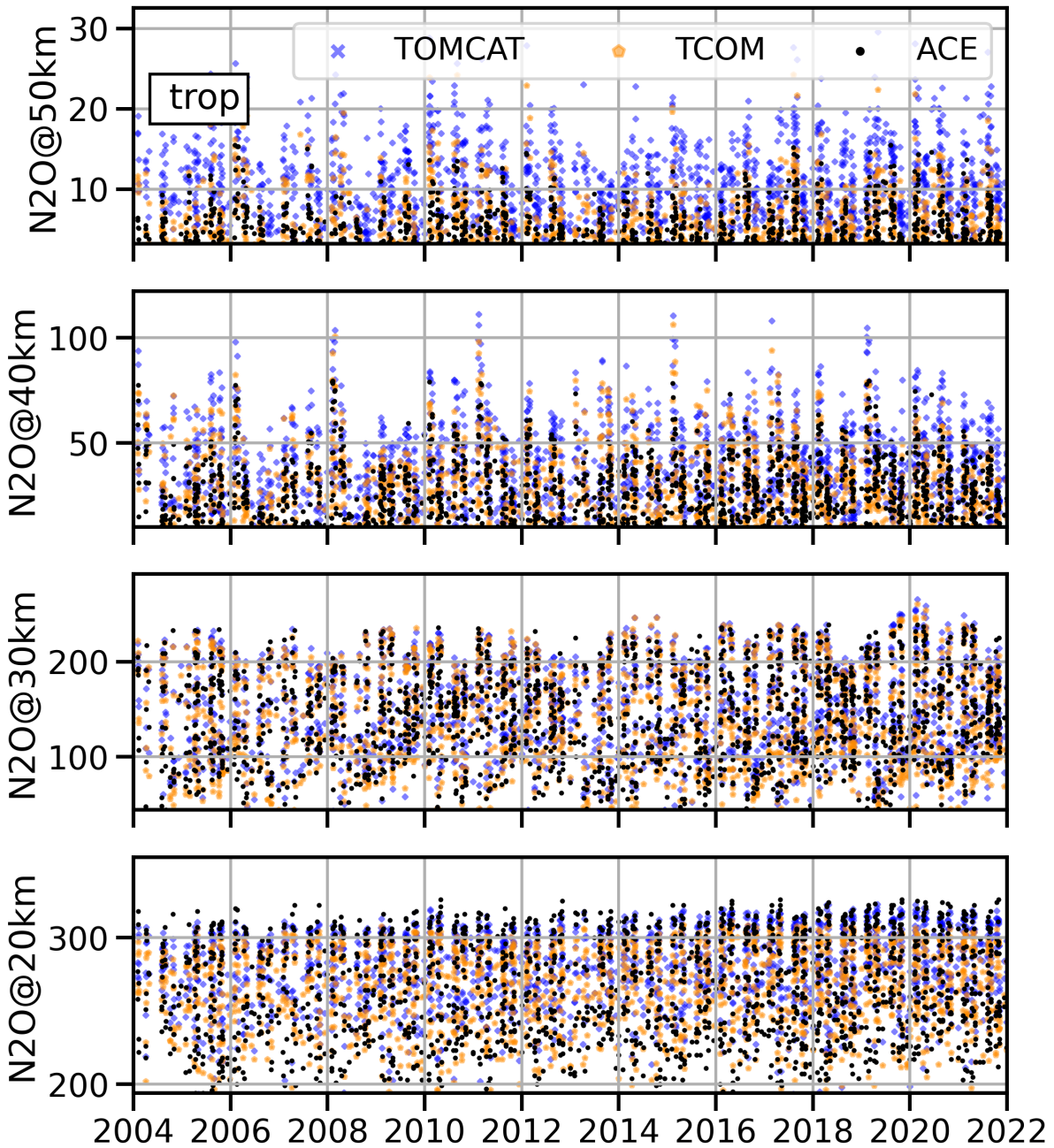
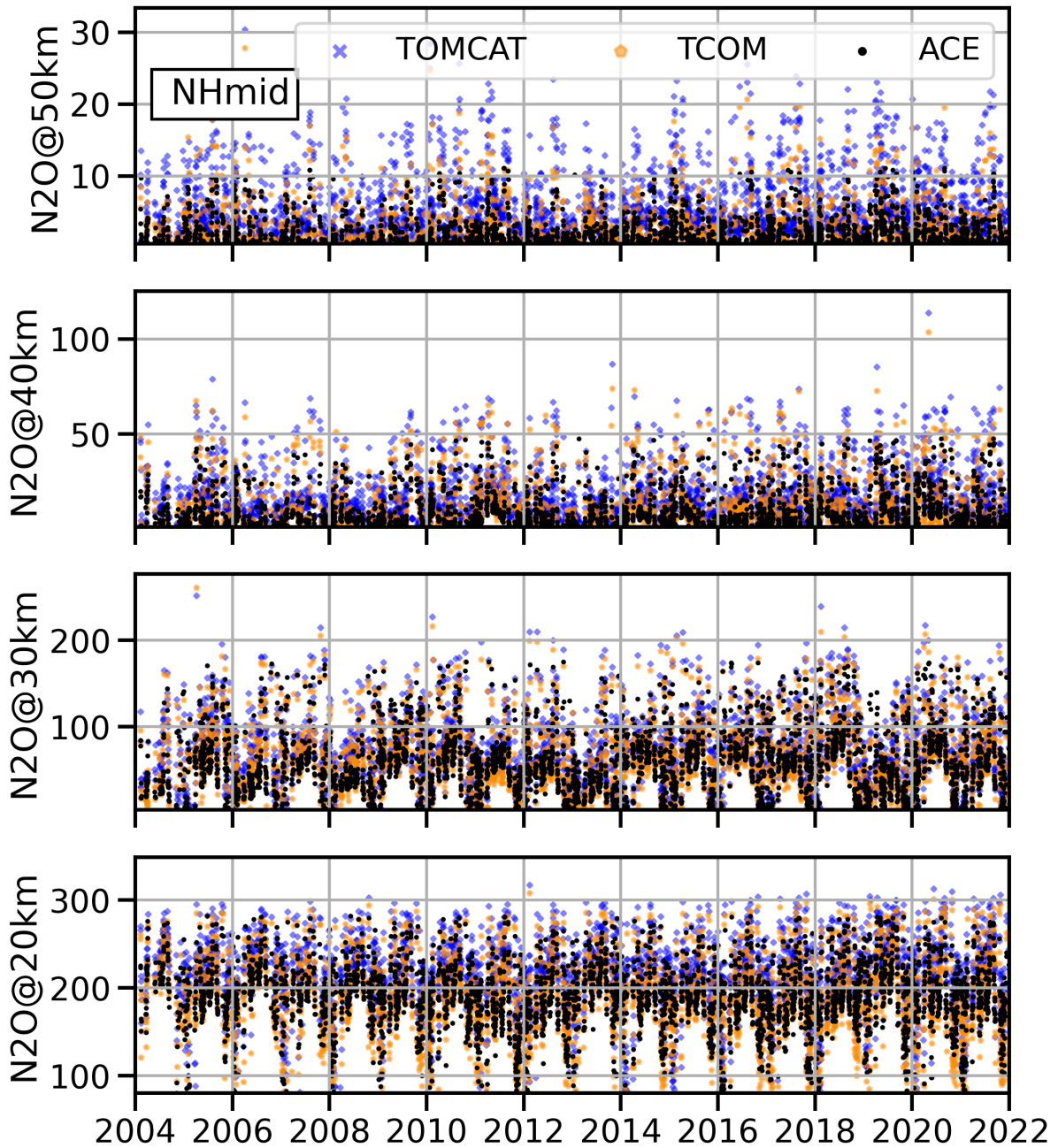**Figure S18.** Same as S17, but for tropical (trop) latitude band (40°S–40°N)

.

Figure S19. Same as S17, but for northern hemisphere mid-latitude (NHmid) band (20°N–70°N)
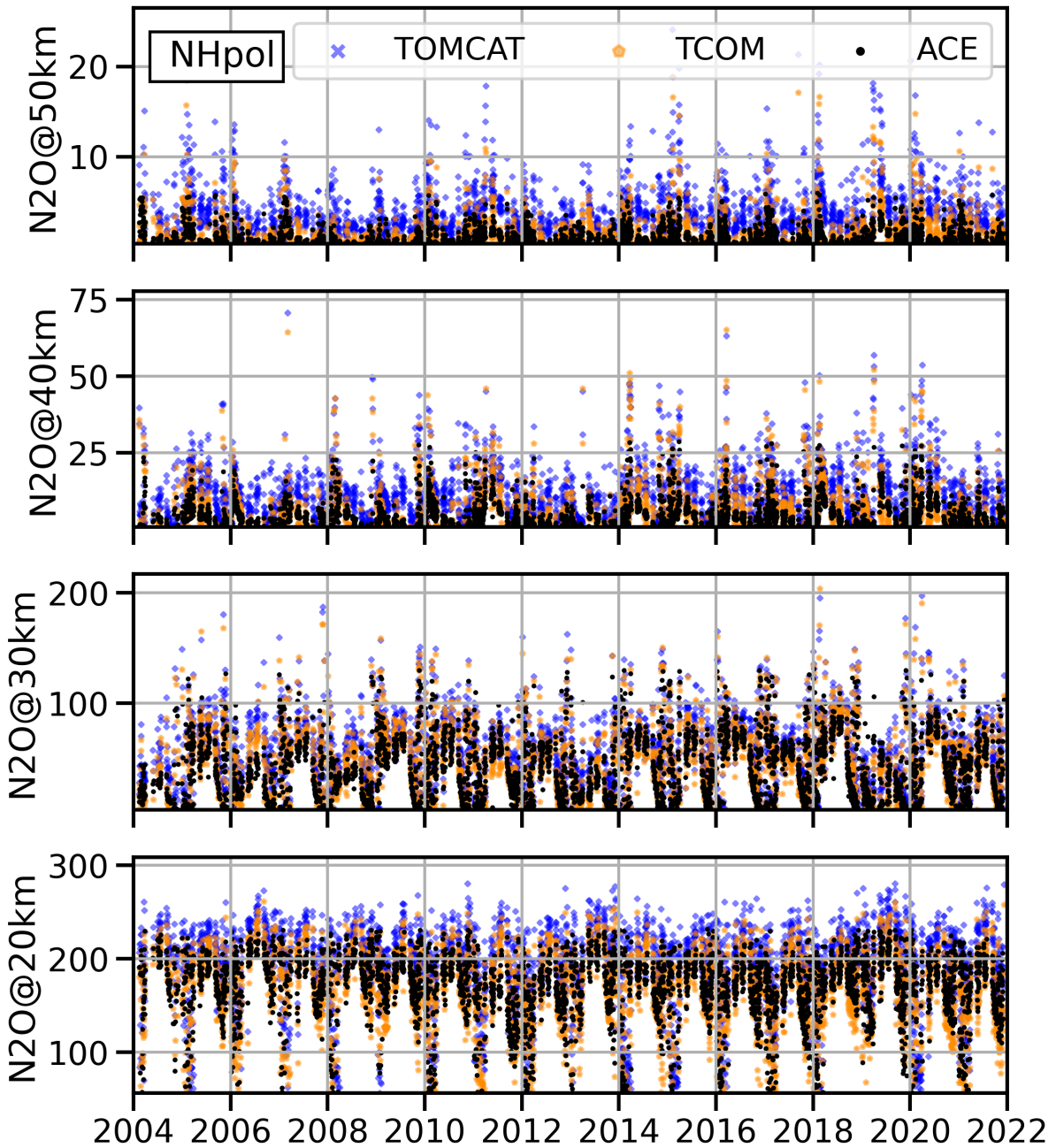
.

Figure S20. Same as S17, but for northern hemisphere polar-latitude (NHpol) band (50°N–90°N)

.