Supplementary Information

S1.1 Sampling and analytical method details

S1.1.1 Air sampling

Research and passenger aircraft sampling and analysis for mixing ratios of halogenated trace gases have been described in previous papers (Kaiser et al., 2006, Laube et al., 2010, Laube et al., 2013, Oram et al., 2017, Leedham Elvidge et al., 2018) including quality assurance and calibration. For the research aircraft, only data from the campaign in 2017 have not been published yet, but that campaign was described in Höpfner et al., 2019. Passenger aircraft data has not been published yet for these species, but the system has been repeatedly proven to perform well for a variety of gases (Laube et al., 2010, 2012, Leedham Elvidge et al., 2015, Oram et al., 2017). More details on all aircraft-based data including sampling coordinates, mixing ratios, and uncertainties can be found in the supplementary data files. All gases here are reported on the most recent NOAA calibration scales, including the recently updated 2016 scale for CFC-11.

The AirCore data is based on the originally published passive sampling method (Karion et al., 2010) and originates from 15 balloon flights. Eleven of these were launched from an established site near Sodankyla in northern Finland (67° 25′ N, 26° 36′ E, Kivi and Heikkinen, 2016), and four near Cambridge in the UK (52.254 °N 0.075 °W). Different AirCores were used, all of which focused on maximising the amount of air collected in the stratosphere. A special version was especially developed for maximising the altitude resolution of the air collected in that region. For that purpose the part of the AirCore that samples tropospheric air and normally consists of ¼" tubing was replaced with ½" tubing, i.e. increasing the internal volume collected into the ⅛" part. All sampled air was dried using a $Mg(ClO_4)_2$-filled trap at the AirCore inlet. More details for these flights are given in Table S1.

Table S1. Details of AirCore flights in Sodankyla, Finland (S) and Elsworth, UK (E). On 20/06/2018 two AirCores were flown on the same payload. Indirect samples (I) were extracted after spectroscopic analysis of the air with a Picarro G2401 analyser (Mrozek et al., 2016), whereas direct sampling (D) translates to a refilling of AirCore air into simple (SSS) and improved (ISS, see text for details) sub-samplers after recovery. In all cases sub-sampler air was later analysed for ppt-level trace gases using the GC-MS system described in the Methods section.

Flight date	Location	AirCore dimensions (length in m * OD in “)	Sampling method	No. of sub-samples	Altitude range (km)
14/04/16	S	40 m * ¼” + 60 m * ⅛”	I-SSS	8	12-28
24/08/16	S	40 m * ¼” + 60 m * ⅛”	I-SSS	9	13-30
21/04/17	S	40 m * ¼” + 60 m * ⅛”	I-SSS	13	8-25
26/04/17	S	40 m * ¼” + 60 m * ⅛”	I-SSS	8	12-27
05/09/17	S	8.5 m * ½” + 63 m * ⅛”	I-ISS	14	11-26
06/09/17	S	8.5 m * ½” + 85 m * ⅛”	I-SSS	8	13-21
07/09/17	S	8.5 m * ½” + 85 m * ⅛”	I-SSS	9	12-23
17/04/18	S	40 m * ¼” + 60 m * ⅛”	I-SSS	14	9-28
01/05/18	E	8.5 m * ½” + 63 m * ⅛”	D-ISS	14	12-30
14/05/18	E	8.5 m * ½” + 63 m * ⅛”	D-ISS	14	12-30
18/06/18	S	40 m * ¼” + 60 m * ⅛”	I-SSS	14	9-24
20/06/18	S	40 m * ¼” + 60 m * ⅛” & 8.5 m * ½” + 63 m * ⅛”	I-ISS (2x)	12+12	12-28, 11-24
27/09/18	E	8.5 m * ½” + 63 m * ⅛”	D-ISS	13	11-22
10/10/18	E	8.5 m * ½” + 85 m * ⅛”	D-ISS	12	13-23

Two different methods were used for extracting the air from the AirCore for analysis. For all Finland-based flights the air was first pushed through a spectroscopic analyser to determine mole fractions of CO_2, CH_4, and CO (as in Karion et al., 2010), and then collected at the exhaust of the pump into a long piece of tubing segmented into 25-50 ml volumes through three-way valves (Mrozek et al., 2016; Paul et al., 2016). The pump was modified to remove small internal leakages. As the AirCore samples were analysed at a constant flow rate through the Picarro analyser, a pre-determined lag time between the internal cell and the sub-sampler was used to determine the start of the AirCore profile and therefore the sub-sampling process.

The second method is based on a 32-port 1/8" valve (from VICI, Switzerland), with a common in- and outlet. We attached loops of ¼" stainless steel to each pair of ports, resulting in 15 loops with an internal volume of about 20 ml each (two ports were blanked for a default position that can be exposed to lab air when connecting the AirCore). All metal surfaces were Silco-1000-treated, which is an established technique to increase the inertness of such surfaces. These new samplers were evacuated directly after connecting the AirCore (having recovered it after a flight) and then the stratospheric end of the AirCore was opened to each loop successively. The tropospheric end of the AirCore was opened to room air to ensure that each loop was filled to the same pressure (and the latter was monitored with a pressure sensor). As the tropospheric part of the AirCore was not sampled and diffusion through the long thin ⅛" tubing of the AirCore is very slow, there was no contamination with lab air.

For both methods diffusion inside the AirCore is negligible over the period of 2-6 hours in between landing and sub-sampling.

S1.1.2 Ensuring contamination-free sampling and measurements

The two most important challenges are to ensure that a) the air is not contaminated throughout the entire sampling and sub-sampling process and b) the precisions of the AirCore samples (15-25 ml) during subsequent GC-MS analysis are comparable to those achieved with the much larger amounts used for aircraft-based samples (200-300 ml). Contaminations can arise from leakages and/or halocarbon-emitting materials (such as organic polymers) in the AirCore itself, in the CO_2-analyser system including the pump, or in the subsampling system. Importantly, for all compounds reported here mixing ratios in the stratosphere are much lower than in even remote tropospheric regions, let alone near sources of these gases. In addition, almost all of the contamination possibilities would affect the entire profile as an AirCore is essentially one air sample. This would become apparent in the correlations of the species with each other, which are very compact in the stratosphere. Figures S1-S4 show correlations of H-1211, H-1301, HCFC-22, and SF_6 with CFC-11 for all aircraft and balloon flights. Note that not all gases were analysed on all flights, and that the research aircraft data (2009-2017), passenger aircraft data (2009-2016), and AirCore data (2016-2018) are grouped together for simplification. HCFC-22 data illustrates what a contamination would look like, as for two AirCore flights (highlighted in red in Figure S3) we did indeed find slightly enhanced mixing ratios resulting in a weaker correlation. As can be seen in Figures 2, S1, S2, and S4, HCFC-22

is the only species that is contaminated for those flights. CFC-11 trends at different mean ages as presented in the main manuscript were also recalculated without using these two flights, but changes to the slopes were much smaller (6-16 times) than the one σ uncertainties of those slopes. We conclude that even if there were undetectable levels of contaminations for the other five gases this would not affect the relevant findings. Several other flights in 2016, 2017 and 2018 showed distinct signs of contamination with non-stratospheric air, either because a different (leakier) Picarro analyser pump was used, or in one case, because an AirCore had developed a leak in the connection between its two lengths of tubing during the flight. Results from these flights were excluded completely.

In addition, if a contamination occurs, one should also observe enhancements of mixing ratios near the tropopause that are well above the northern hemispheric background values at that time. Figures S5 and S6 show these enhancements for five of the gases (excluding H-1301 due to the lack of an available northern hemispheric time series and the limited measurements precisions). Only small enhancements are observed, both from aircraft as well as from AirCores and there are almost no mixing ratios enhanced above the northern hemispheric background by more than two measurement standard deviations, which gives further confidence in the data.

Finally, all AirCores, sub-samplers and connections were leak-checked before each flight/filling. AirCores were conditioned and filled with nearly trace gas-free fill gas (either air or nitrogen; low ppt-level mixing ratios observed for some gases) before each flight. Since the uppermost one or two samples still contain a portion of this fill gas and correcting for that is complicated and introduces uncertainties, these samples were excluded for the analysis in this work. Sub-samplers were conditioned with either fill gas (Finland flights) or ultra-clean Helium (UK flights) before filling them with AirCore air. Storage tests of unpolluted air in AirCores and sub-samplers showed, within one σ uncertainties, no effect on any of the six gases reported here after several weeks.

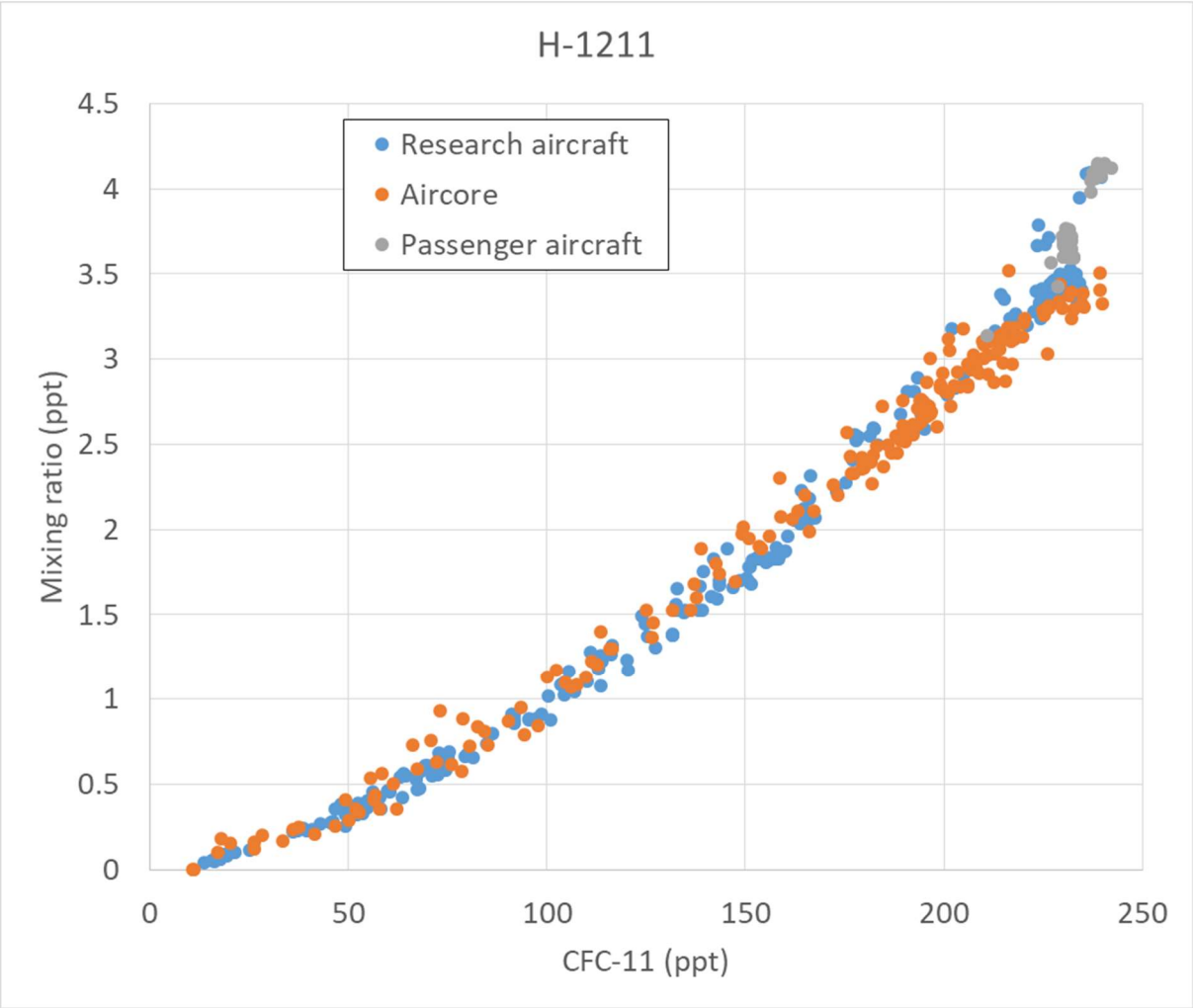

Figure S1. Stratospheric and upper tropospheric correlation of H-1211 mixing ratios with those of CFC-11. Tropospheric mixing ratios of H-1211 have substantially decreased during 2009-2018 and the data near the tropopause (top right) follows that trend well as is also shown in Figure 1.

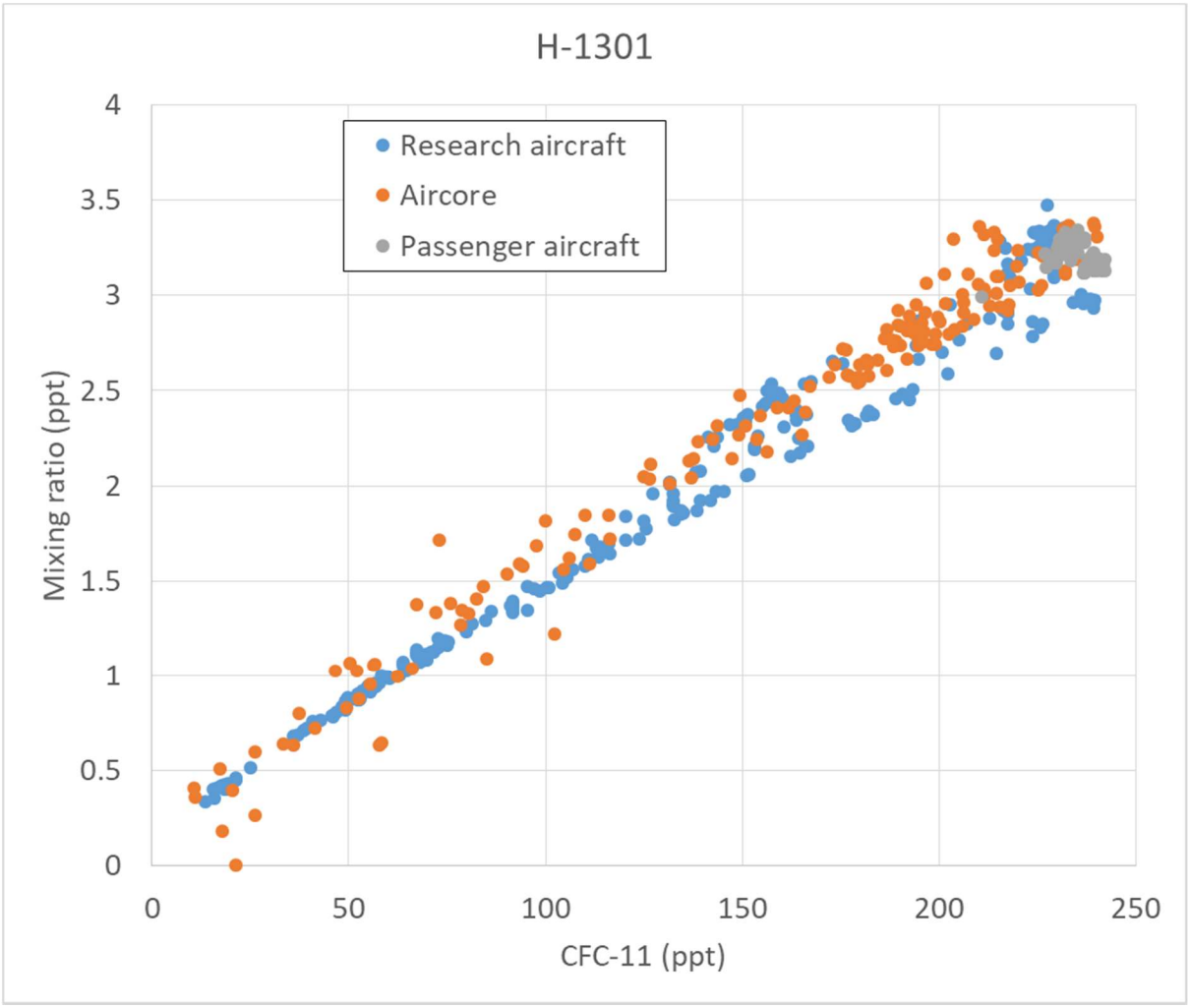

Figure S2. Stratospheric and upper tropospheric correlation of H-1301 mixing ratios with those of CFC-11. The larger variability of the AirCore data in the deeper part of the stratosphere (bottom left) is caused by worse measurement precisions as the signal approaches detection limits (Table S2).

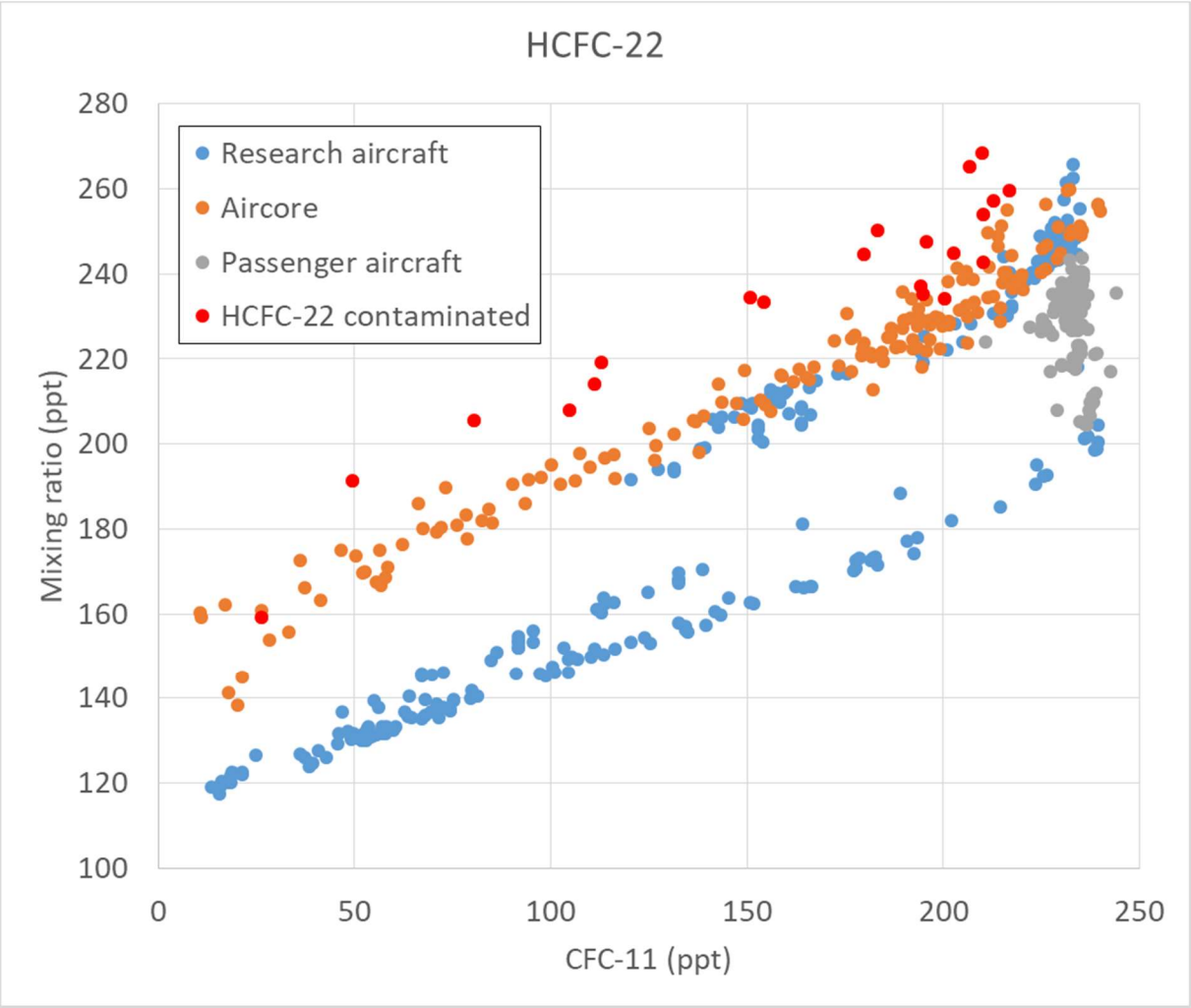

Figure S3. Stratospheric and upper tropospheric correlation of HCFC-22 mixing ratios with those of CFC-11. Tropospheric HCFC-22 mixing ratios have increased continuously between 2009 and 2018 and the shifts in the correlations reflect that. Two AirCore flights showing a slight contamination of the profile with HCFC-22 are highlighted in red and were not included in Figure 1.

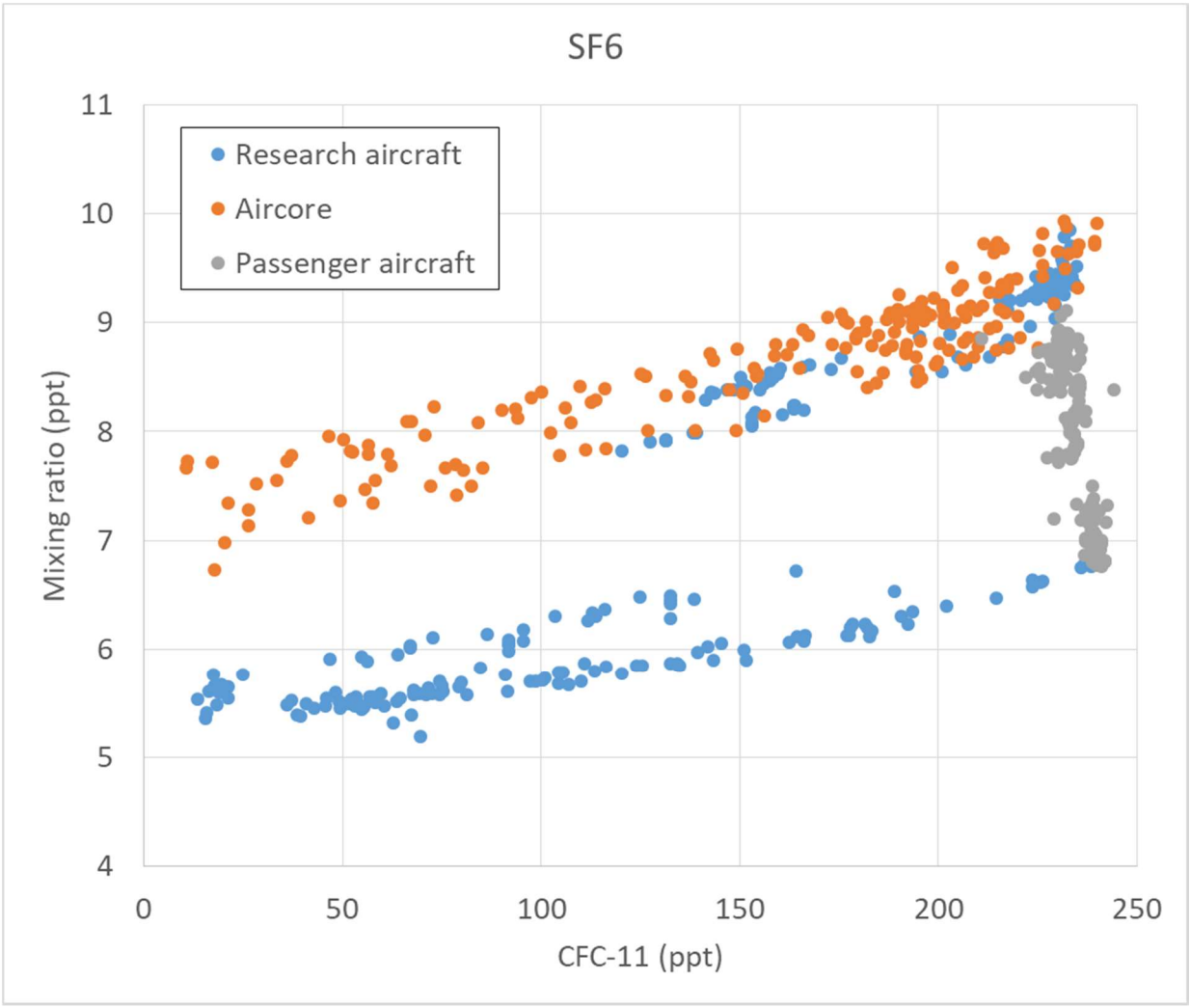

Figure S4. Stratospheric and upper tropospheric correlation of SF_6 mixing ratios with those of CFC-11. Tropospheric SF_6 mixing ratios have increased continuously between 2009 and 2018 and the shifts in the correlations reflect that.

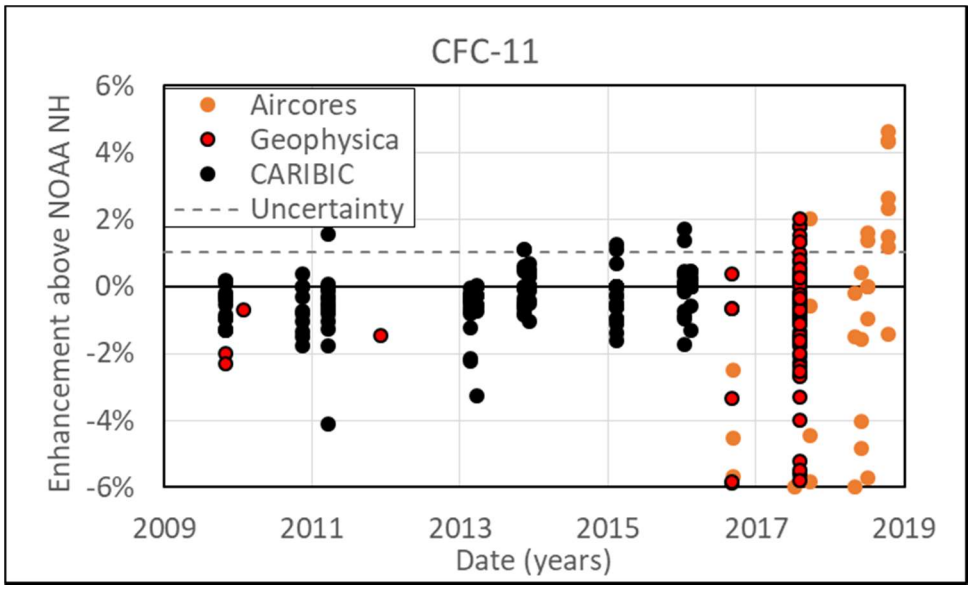

Figure S5. Enhancements of CFC-11 mixing ratios near the tropopause above the NOAA northern hemispheric background. The dashed line is the sum of the average NOAA measurement uncertainty over the period shown and that of the AirCore measurements with the latter derived as one standard deviation of the repeated working standard measurements (see Table S2 for further details).

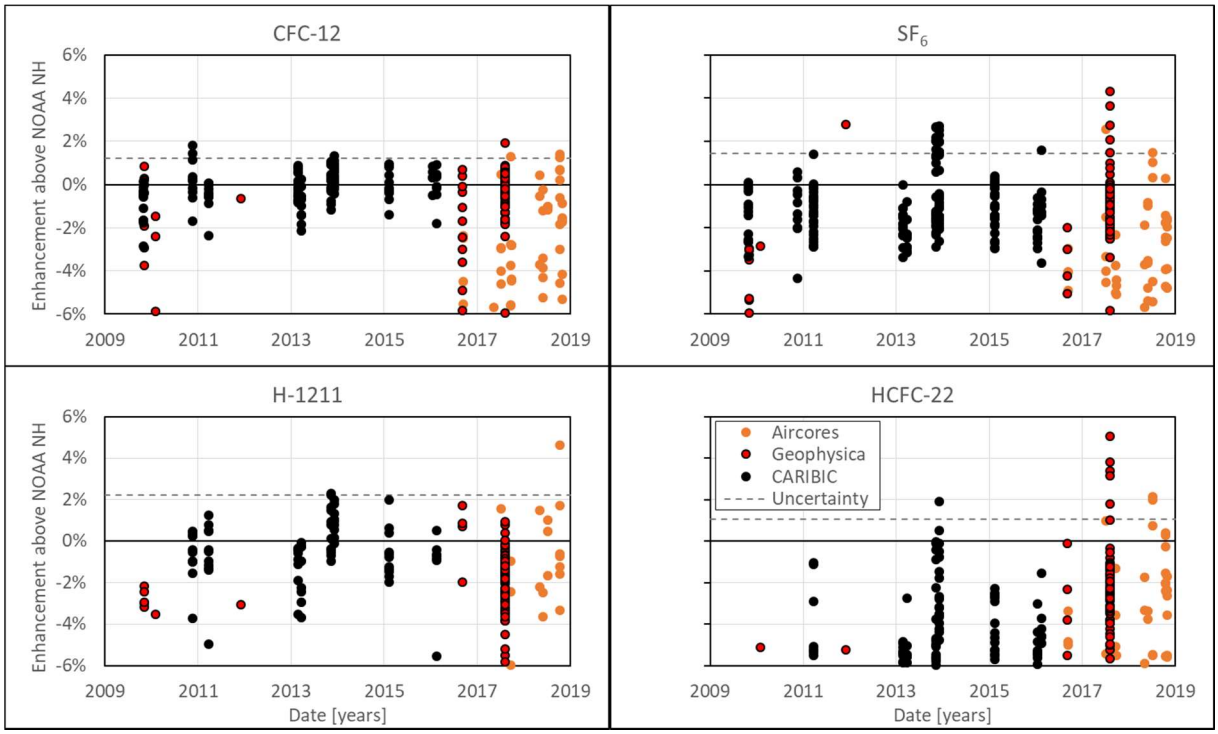

Figure S6. The same as in Figure S5 but for CFC-12, H-1211, HCFC-22, and SF_6. For HCFC-22 the NOAA time series from Mace Head, Ireland was used as it is geographically closest to most aircraft and balloon launch sites (most of which are in Europe).

Table S2. Comparison of average measurement uncertainties (derived as the average one standard deviation from repeated working standard or air sample measurements) of the research aircraft campaign in 2016, all AirCore flights, and some AirCore sample repeats for CFC-11, CFC-12, H-1211, H-1301, HCFC-22, and SF_6. For AirCore uncertainties the average working standard uncertainty over three years was used as it is a) more representative of the entire measurement period and b) generally comparable or worse than precisions derived from sample repeats. AirCore-based precisions are generally slightly worse than those achieved with the larger aircraft-based samples, but still much smaller than mixing ratio gradients observed in the stratosphere.

Trace Gas / Average precision [%]	Aircraft 2016	AirCore 2016-18 Standards	AirCore 2017 Sample repeats
CFC-11 ($CFCl_3$)	0.4	0.9	1.2
CFC-12 (CF_2Cl_2)	1.1	1.2	0.8
H-1211 (CF_2ClBr)	0.6	1.9	1.0
H-1301 (CF_3Br)	0.6	3.3	2.3
HCFC-22 (CHF_2Cl)	0.6	0.9	0.2
SF_6	0.4	0.9	0.6

S1.1.3 Assigning coordinates to AirCore samples

As has been pointed out before, one of the more complicated aspects of AirCore-based trace gas measurements is the reliable assignment of altitudes (and, to a lesser degree, latitudes and longitudes) to individual samples (Karion et al., 2010, Membrive et al., 2017, Engel et al., 2017). This is mostly caused by a disequilibrium between the inside of the long narrow AirCore tubing and the outside air, especially in the stratospheric part of the profile, when the

payload is descending at high speeds. However, as opposed to the optical spectroscopy-based measurements we here average over air sections of 20 or more ml due to the sub-sampling process. This limits the altitudinal resolution and therefore also the influence of the aforementioned disequilibrium. We derived a simple algorithm to assign altitudes, which uses the minimum pressure measured by the radiosonde that accompanies each AirCore as the top of the profile. For simulating the AirCore filling process during descent we then apply the ideal gas equation to calculate the number of moles of air sampled from the radiosonde pressure, the AirCore volume, and the temperature of the AirCore itself as recorded with individual PT100-type temperature sensors attached to the AirCore tubing. These are then used, in combination with laboratory temperatures and pressure to calculate the volume that was transferred into the subsampler. A comparison of retrieval algorithms between our approach and the NOAA-based one (developed from Karion et al., 2010) for the flights on 5th and 6th September 2017 showed agreement of the derived altitudes within 360 meters. To put this into context, the uppermost two air samples subsampled from those two flights were 1.4 and 3.3 km apart, and air in subsamples at those altitudes is averaged over up to 4 km, so the retrieval does not have a major influence on our sample altitudes.

We note that, while improving these retrievals is certainly desirable, we here use the derived altitudes, latitudes and longitudes only for sampling the model data at the derived observation coordinates for the comparison in Figure 2. The average CFC-11 mixing ratio difference observed (observations-model, AirCore data only) was -8.4 ppt for CLaMS-ERA-Interim, 8.7 ppt for CLaMS-JRA-55, and 10.7 ppt for CLaMS-MERRA-2. AirCore coordinates do not affect the comparisons with ground-based records nor the correlation-based analysis of CFC-11 and CFC-12 at certain mean ages.

S1.2 Further details on determining temporal trends of CFC-11 and CFC-12 at different mean ages of air

Aircraft-based AoA was calculated, where available, from two to five gases (i.e. C_2F_6, C_3F_8, HFC-23, HFC-125, and CF_4) that have recently been introduced and evaluated as new age tracers as out lined in Ray et al., 2017 and Leedham Elvidge et al., 2018. For the AirCore data, only measurements of SF_6 were available at sufficient precisions. As demonstrated in Leedham Elvidge et al., 2018 (and previously also in Andrews et al., 2001 when comparing with CO_2-based AoAs) we correct for a high-AoA bias from SF_6 by multiplying all SF_6-

derived AoAs with a conversion factor of 0.85. This leads to a much better agreement of the CFC-11-AoA correlations with those derived from aircraft. Not using such a correction factor would also lead to larger CFC-11 mixing ratios at certain mean ages and would therefore produce even more positive temporal trends in the stratosphere. We recognise that this introduces an additional and not quantifiable uncertainty to the AirCore-based AoAs and have therefore referred to this in the conclusions. It should also be noted that the AoA calculation from observations makes use of a parametrisation (based on Engel et al., 2002) of the so-called age spectrum, i.e. the distribution of transit times underlying each mean age. This represents an unquantifiable uncertainty as any dynamical changes in the stratosphere will inherently change this distribution. However, as shown in Leedham Elvidge et al., 2018, it is encouraging that with this parameterisation a range of different trace gases produce consistent AoA estimates from various aircraft campaigns, despite their very different tropospheric time series used for such calculations. In addition, changes of the parameterisation itself were assessed in Leedham Elvidge et al., 2018, and found to play only a minor role in the estimated mean age uncertainties.

Table S3 shows the slopes derived from observational and model data for CFC-11, CFC-12 – and for observations only, HCFC-22 and H-1211 - between 2009 and 2018 at four different mean ages of air. Model data from all three reanalyses consistently produces more negative trends for both gases than observation-based data at mean ages of one to three years, i.e. in the lower part of the stratosphere, where most of its mass resides (see also Figures 3 and S7-S13). In addition, the three runs seem to overestimate the mixing ratios of CFC-11 and CFC-12 present in the lower stratosphere (one and two years AoA), while underestimating it at an AoA of 4 years – apart from MERRA-2 which consistently overestimates at all AoAs and has the highest offsets to observations.

Table S3. Temporal trends and their two σ uncertainties of CFC-11 and CFC-12 mixing ratios at AoAs of one, two, three, and four years. These slopes correspond to an uncertainty-weighted regression line fitted to the data in Figures 3, and S7-S9, with two exceptions: 1) The data from the Asian Monsoon campaign in 2017 was excluded as this is region is not representative of NH stratospheric air and 2) all data at mean ages above 3.5 years from winter campaigns in high latitudes was also excluded as it might contain polar vortex air, which is equally unrepresentative. Model-based slopes were derived over the same period as

observational data (2009.8-2018.8), except for JRA-55 and MERRA-2, where data was only available until the end of 2017.

CFC-11/AoA	1 year	2 years	3 years	4 years
Slope obs (ppt/year)	0.69	1.77	1.25	0.59
Uncertainty (ppt/year)	1.54	1.81	1.60	2.12
Trend (%/decade)	3.2	10.4	10.2	7.4
Slope ERA-I	-1.35	-0.50	1.15	3.09
Uncertainty	0.22	0.24	0.47	0.61
Slope JRA-55	-1.56	-1.38	-0.08	1.73
Uncertainty	0.21	0.20	0.27	0.62
Slope MERRA-2	-1.69	-1.51	-1.20	-0.55
Uncertainty	0.18	0.23	0.23	0.30

CFC-12				
Slope obs (ppt/year)	-1.96	-0.45	-0.38	-1.33
Uncertainty (ppt/year)	1.90	2.20	2.52	5.36
Trend (%/decade)	-3.6	-0.95	-0.93	-3.9
Slope ERA-I	-3.09	-2.52	-1.52	3.21
Uncertainty	0.37	0.48	0.62	1.20
Slope JRA-55	-3.17	-3.09	-1.37	2.39
Uncertainty	0.28	0.41	0.60	1.08
Slope MERRA-2	-3.26	-3.17	-3.02	-2.40
Uncertainty	0.24	0.39	0.51	0.76

HCFC-22				
Slope obs (ppt/year)	6.15	6.16	5.98	5.67
Uncertainty (ppt/year)	0.15	0.14	0.14	0.18
Trend (%/decade)	30.5	33.4	36.0	38.2

H-1211				
Slope obs (ppt/year)	-0.031	0.000	0.013	0.013
Uncertainty (ppt/year)	0.008	0.008	0.007	0.009
Trend (%/decade)	-9.0	0.2	9.1	22.4

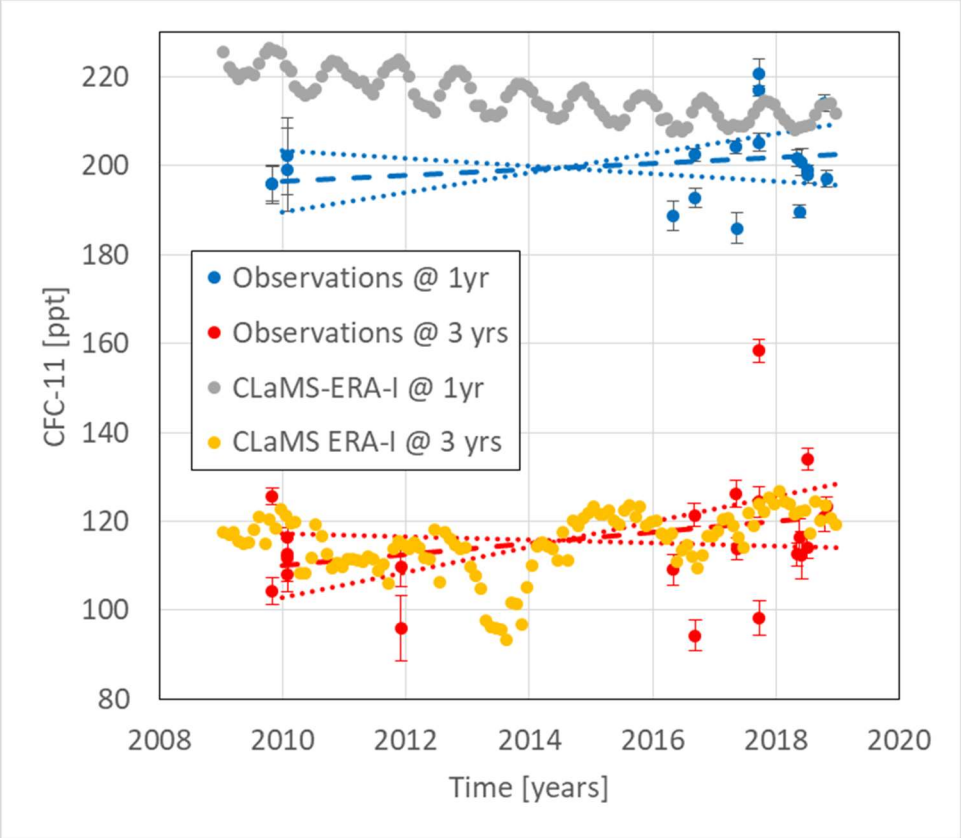

Figure S7. The same as in Figure 3 but at AoAs of one and three years.

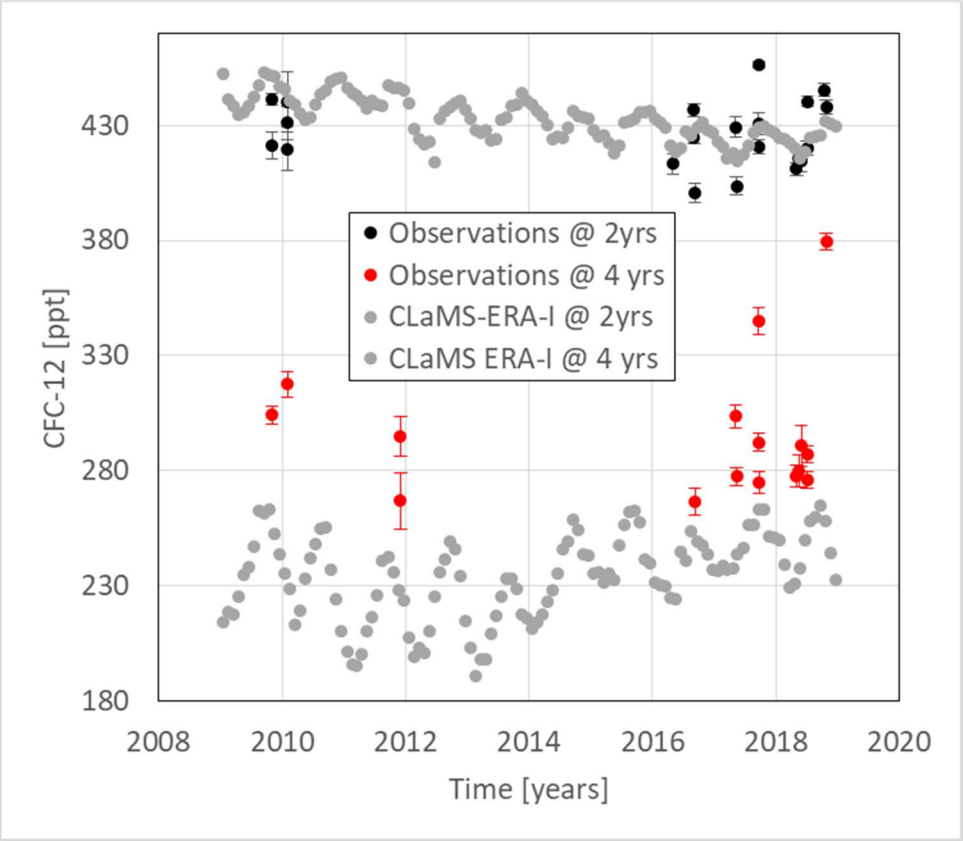

Figure S8. The same as in Figure 3 but for CFC-12.

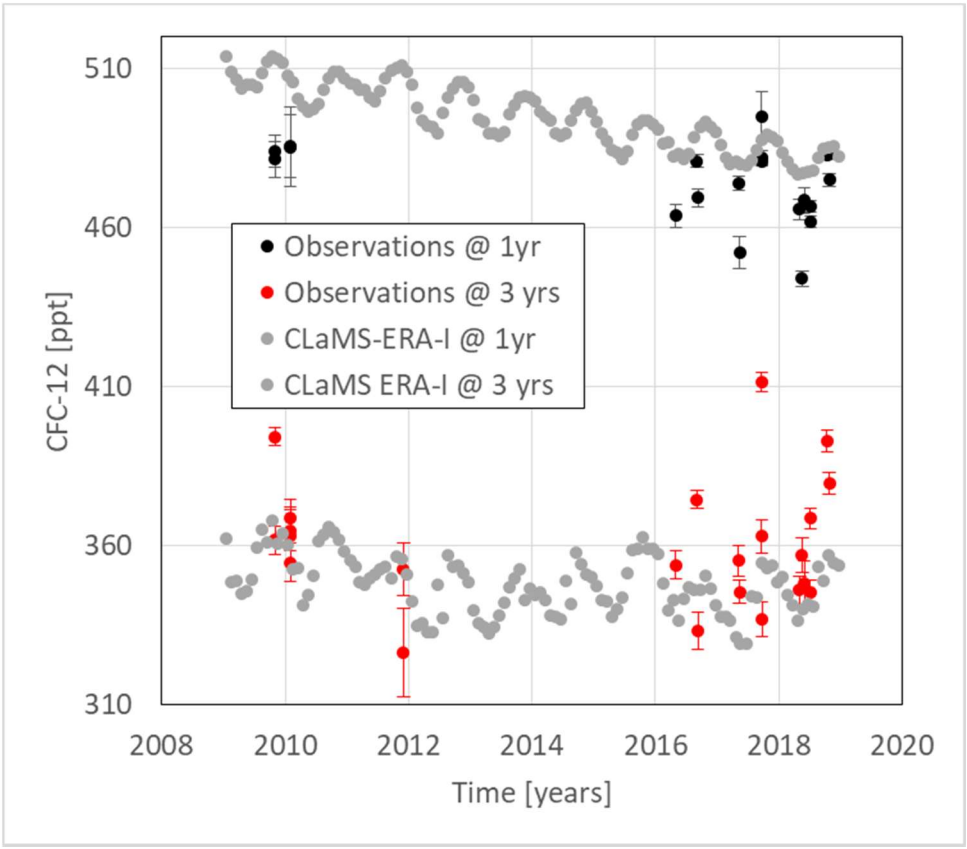

Figure S9. The same as in Figure 3 but for CFC-12 and at AoAs of one and three years.

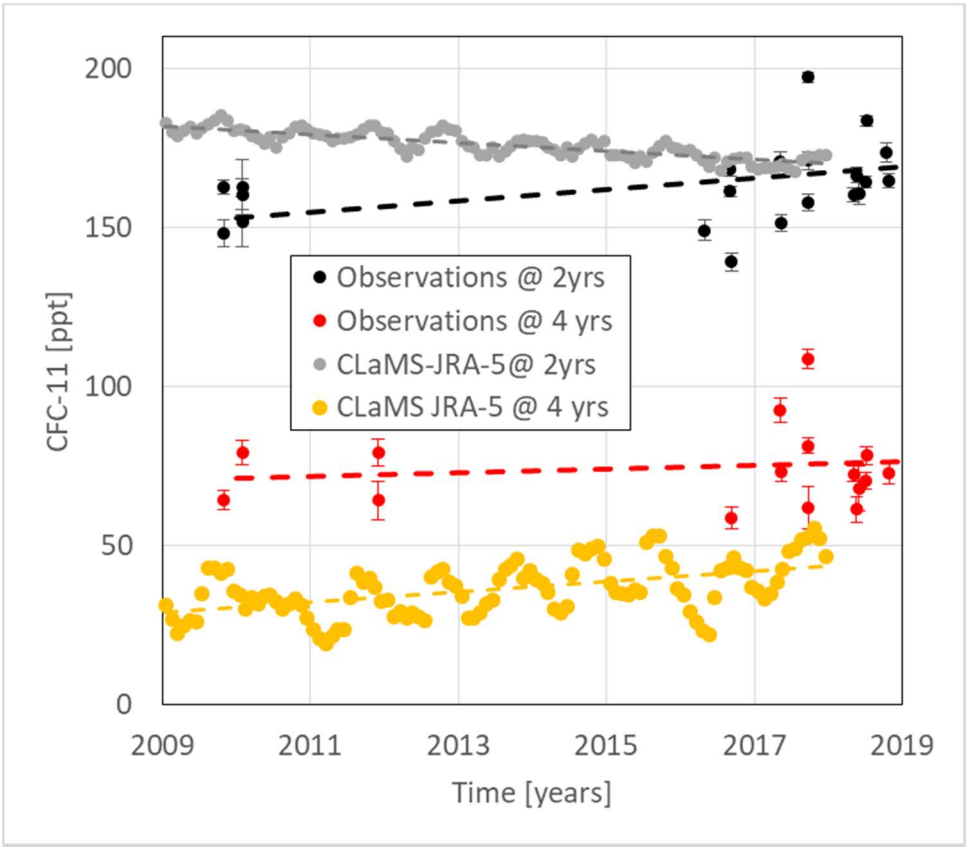

Figure S10. The same as in Figure 3 but for CLaMS-JRA-55.

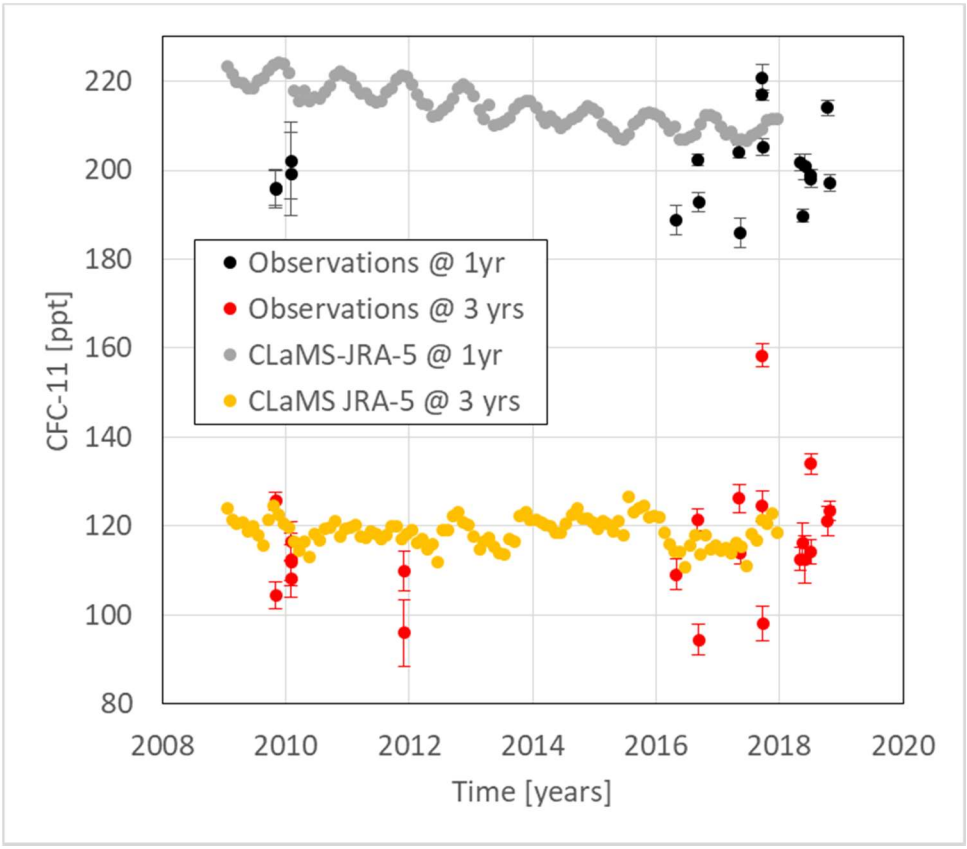

Figure S11. The same as in Figure 3 but for CLaMS-JRA-55 and at AoAs of one and three years.

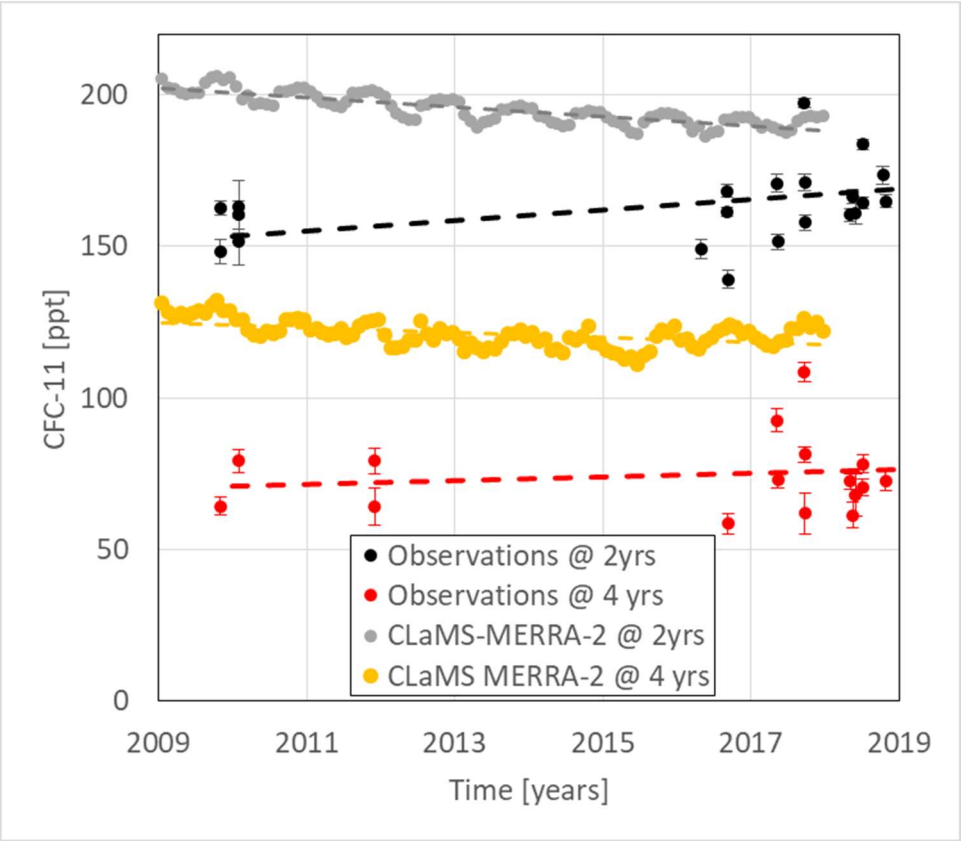

Figure S12. The same as in Figure 3 but for CLaMS-JRA-55.

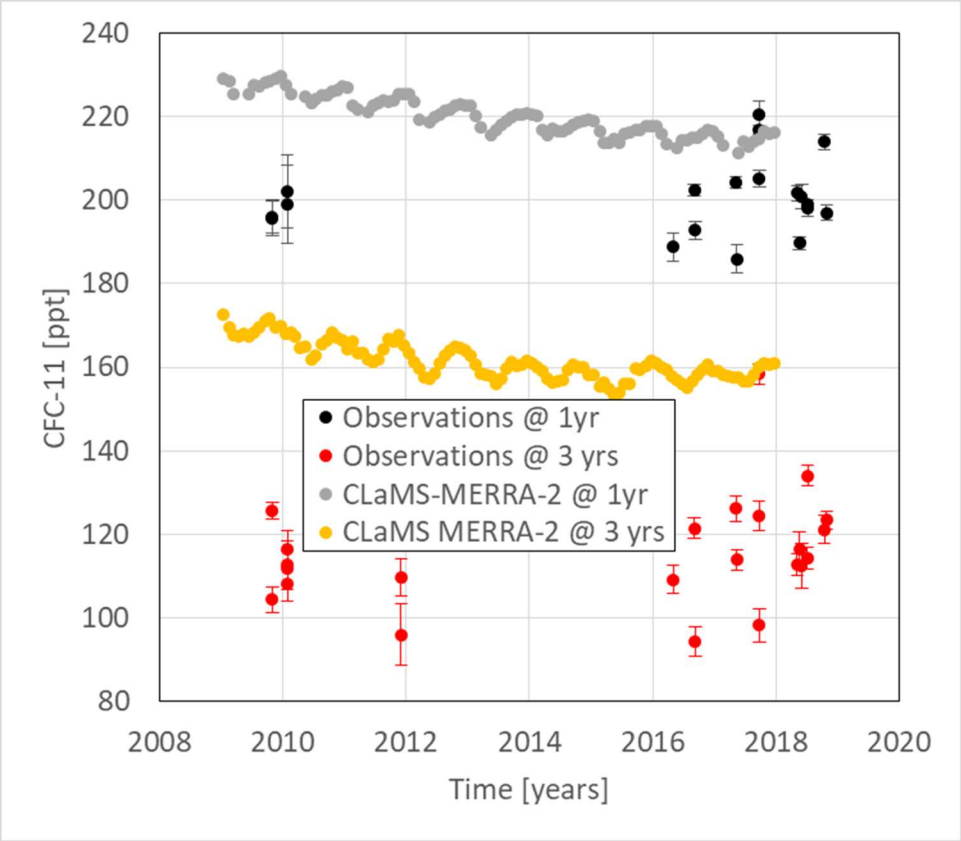

Figure S13. The same as in Figure 3 but for CLaMS-JRA-55 and at AoAs of one and three years.

S1.3 Further details on the CLaMS model runs

The Chemical Lagrangian Model of the Stratosphere (CLaMS) is a Lagrangian chemical transport model, with advective transport calculated from three-dimensional forward trajectories and an additional parameterisation for small-scale turbulent mixing (McKenna et al., 2002). Potential temperature is used as vertical coordinate throughout the stratosphere with vertical velocity estimated from the total diabatic heating rate. Further model details and the chemistry scheme used are described in Pommrich et al. (2014). For the simulations used in this study CLaMS was driven with horizontal winds and diabatic heating rates from three alternative meteorological reanalysis data sets: ERA-Interim (from European Centre for Medium-Range Weather Forecasts, ECMWF), JRA-55 (from Japan Meteorological Agency), and MERRA-2 (from NASA).

The stratospheric Brewer-Dobson circulation and AoA in these three reanalyses have been recently compared by Chabrillat et al. (2018) and Ploeger et al. (2019). The mean age, CFC-11 and CFC-12 changes around the year 2013, which are most relevant for the analysis presented here, are shown in Fig. S15. Different definitions of the tropopause were tested for the CFC-11 mass flux estimates (Figure 4), i.e. the 3 PVU surface as well as the tropopause as defined by the WMO. The results for the corresponding CFC-11 mass flux change around 2013 were very similar i.e. within a few percent of each other.

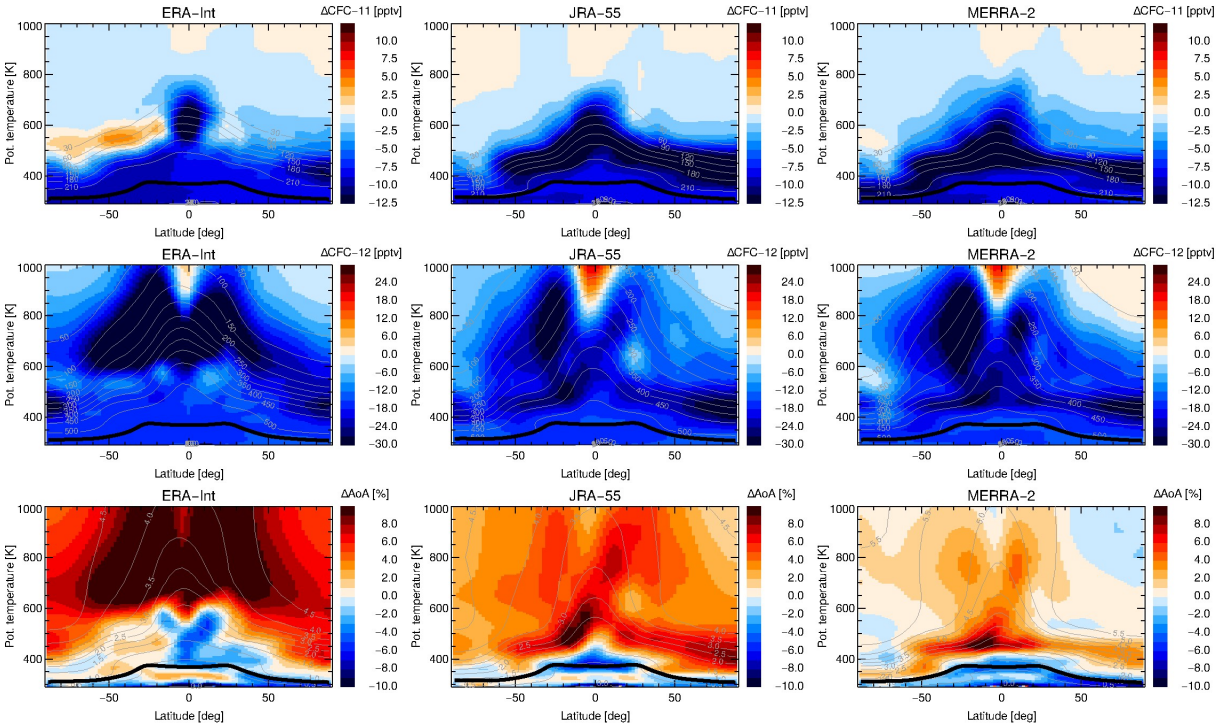

Figure S15. Stratospheric CFC-11, CFC-12, and AoA changes (in percent) around the year 2013 of three CLaMS runs driven by ERA-Interim, JRA-55, and MERRA-2. Changes are calculated as differences between the later period 2013-2017 and the earlier period 2008-2012. The grey lines show climatological CFC-11, CFC-12 and AoA contours (for 2008-2012), the thick black line the tropopause.

Supplement References

1. Kaiser, J. et al. Probing stratospheric transport and chemistry with new balloon and aircraft observations of the meridional and vertical N_2O isotope distribution. *Atmos. Chem. Phys.* **6**, 3535–355 (2006).

2. Kivi, R. and Heikkinen, P. Fourier transform spectrometer measurements of column CO_2 at Sodankylä, Finland. *Geosci. Instrum. Method. Data Syst.* **5**, 271-279 (2016).

3. Andrews, A. E. et al. Mean ages of stratospheric air derived from in situ observations of CO_2, CH_4, and N_2O. *J. Geophys. Res.* **106**, 32295–32314 (2001).

4. Engel, A. et al. Temporal development of total chlorine in the high-latitude stratosphere based on reference distributions of mean age derived from CO_2 and SF_6. *J. Geophys. Res.-Atmos.* **107**, ACH 1-1–ACH 1-11 (2002).

5. Chabrillat, S. et al. Comparison of mean age of air in five reanalyses using the BASCOE transport model. *Atmos. Chem. Phys.* **18**, 14715–1473 (2018).