Two decades of in-situ temperature measurements in the upper troposphere and lowermost stratosphere from IAGOS long-term routine observation

Florian Berkes[1], Patrick Neis[1], Martin G. Schultz[1], Ulrich Bundke[1], Susanne Rohs[1], Herman G.J. Smit[1], Andreas Wahner[1], Paul Konopka[2], Damien Boulanger[3], Philippe Nédélec[3], Valerie Thouret[3], and Andreas Petzold[1]

[1]Forschungszentrum Jülich, IEK-8, Jülich, Germany
[2]Forschungszentrum Jülich, IEK-7, Jülich, Germany
[3]Laboratoire d'Aérologie, CNRS and Université de Toulouse, Toulouse, France

Correspondence to: Florian Berkes (f.berkes@fz-juelich.de)

Abstract. Despite several studies on temperature trends in the tropopause region, a comprehensive understanding of the evolution of temperatures in this climate-sensitive region of the atmosphere remains elusive. Here we present a unique global-scale, long-term data set of high-resolution in-situ temperature data measured aboard passenger aircraft within the European Research Infrastructure IAGOS (In-service Aircraft for a Global Observing System, www.iagos.org). This data set is used to investigate temperature trends within the global upper troposphere and lowermost stratosphere (UTLS) for the period 1995 to 2012 in different geographical regions and vertical layers of the UTLS. The largest amount of observations is available over the North Atlantic. Here, a neutral temperature trend is found within the lowermost stratosphere. This contradicts the temperature trend in the European Centre for Medium Range Weather Forecast (ECMWF) ERA-Interim reanalysis, where a significant (95% confidence) temperature increase of +0.56 K/decade is ~~obtained~~found. Differences between trends derived from observations and reanalysis data can be traced back to changes in the temperature ~~bias difference~~ between observation and model data over the studied period. This study ~~demonstrates~~ underpins the value of the IAGOS temperature observations as anchor point for the evaluation of reanalyses and its suitability for independent trend analyses.

1 Introduction

Temperature changes in the lower stratosphere (~50 hPa) obtained from radiosondes and satellite retrievals show cooling of about 0.5 K/decade over much of the globe during the period from 1979 to 1995. Since 1995, the cooling turned into a neutral trend with larger increase (not significant) over the Antarctic region than over the tropics (Randel et al., 2009; Blunden et al., 2014; Seidel et al., 2016~~;~~). The robustness of the temperature trends in the lower stratosphere derived from radiosondes (since 1958), and from satellites (since 1979) suffers from instrumental uncertainties such as sensor changes,

drifts, etc., implying large uncertainties in the trend estimates (Simmons et al., 2014). In recent years, several studies assessed the uncertainty of temperature trends in the lower stratosphere, and the impact on changing trends of radiatively active constituents (such as ozone) or ~~the~~ atmospheric dynamics (e.g. Fueglistaler et al., 2014; Seidel et al., 2016).

5   Temperature trends in the upper troposphere - lowermost stratosphere (UTLS) are even more uncertain due to insufficient regional coverage of in-situ observations. Most studies of UTLS temperature trends are based on the global radiosonde network, but most of these data are from the Northern Hemisphere mid-latitudes (70% of radiosonde launches occurred between 30° and 60°N). Furthermore, these observations suffer~~s~~ from time-varying biases, which cannot capture the large variability of the UTLS and inhomogeneity due to changes in instrumentation (Bencherif et al., 2006; Seidel and Randel, 2006; Xu and Powell, 2010). Satellite observations cover the spatial scale, but are limited by their coarse vertical resolution, especially in the UTLS region. A well-suited data source for temperature profiles is the relatively new global positioning system radio occultation technique (GPSRO) (Kursinski et al., 1997; Wickert et al., 2001). The reliability of these measurements within the UTLS region has been demonstrated with trend analyses of GPSRO temperature (Steiner et al., 2009), or GPSRO derived trends of the thermal tropopause temperature and tropopause height (Schmidt et al., 2010; Rieckh et al., 2014; Khandu et al., 2016). Ho et al. (2017) demonstrated the usefulness of the GPSRO measurements to correct the temperature bias of radiosondes with different sensor types in the ~~LMS~~ lowermost stratosphere (LMS), which is an important task to reduce the temperature uncertainties. Comparison of tropospheric temperature trends derived from homogenized satellite data sets and model simulations find more consistency, but require long-term time series (>17 years) before a robust trend arises from internal climate variability (Santer et al., 2011).

20

Since the radiative forcing from greenhouse gases, including water vapor, is sensitive to changes in the mid-troposphere and the UTLS (Solomon et al., 2010; Riese et al., 2012), this region is extremely important for climate change and for controlling dynamical processes governing stratosphere-troposphere exchange (Gettelman et al., 2011). Furthermore, the variability and changes of the temperature in the UTLS play an important role in regulating the exchange of water vapor, ozone, and other trace gases between the troposphere and the stratosphere.

Continuous in-situ observations of these properties in the UTLS region can only be conducted with satellite and aircraft measurements over a large spatial region. Automated aircraft temperature observations are collected, along with an increasing amount of humidity data through the World Meteorological Organization (WMO) Aircraft Meteorological DAta Relay (AMDAR) program (WMO, 2014; Petersen ~~et al.~~, 201~~5~~6). Petersen ~~et al.~~ (201~~6~~5) showed that at flight level the errors of temperature and wind in the 3-48 h forecast were reduced by nearly 50% when ~~including~~ assimilating data from passenger aircraft. However, Ballish and Kumar (2008) and Drüe~~Drue~~ et al. (2008) identified that the AMDAR aircraft temperature is strongly affected by a warm bias, which can fluctuate by altitude, aircraft type and phase of flight, while the reason for this bias is not fully understood (Ingleby et al., 2016).

Previous studies already used the passenger aircraft measurements from MOZAIC (Measurement of Ozone and Water Vapour on Airbus In-service Aircraft, Marenco et al., 1998) temperature measurements for inter-comparison with GPSRO measurements and ECMWF (European Centre for Medium Range Weather Forecast) analyses. Bortz et al. (2006) analyzed

5 MOZAIC temperature measurements at cruise altitude from 1994 to 2003 within the tropical region. They found no significant temperature increase within the upper tropical troposphere. The authors concluded that the temperature measurements are well representative to be used in inter-comparison studies with satellite (MSU) and radiosonde measurements within this region, but at this time the data record was too short for trend estimates. Heise et al. (2008) compared around 2700 MOZAIC in-situ temperature profiles against profiles from GPSRO and analysis from ECMWF

10 between 2001 and 2006. They concluded that MOZAIC in-situ temperature had no bias against the ECMWF temperature above 300 hPa, whereas GPSRO showed a cold bias of -0.9 K compared to the MOZAIC temperature. Since 2001 GPSRO data are assimilated in numerical weather prediction models and reanalysis products. Schmidt et al. (2010) characterized the tropopause inversion layer in the Northern Hemisphere with temperature profiles from in-situ measurements and from the model both types of profile measurements for the period from 2001 to 2009 and concluded that the cold point at the

15 tropopause agreed well.

In this study, IAGOS (In-service Aircraft for a Global Observing System) temperature observations are presentedanalyzed, which are available for almost two decades since 1994. The geographical coverage of the measurements is shown in section 2, where the reliability of the IAGOS observations and the data selection are discussed. The UTLS temperature distribution

20 and derived temperature trends are presented in section 3, and their robustness and suitability for the evaluation of global scale reanalyses with the example of ERA-Interim are discussed in section 4. Due to varying geographical coverage of these data they cannot provide a full global assessment yet, but as we will show in our study, they can serve as anchor point for trend analyses and evaluation of reanalysis at least over the extratropics. Thise conclusion is included in section 5.

2 Data selection and methods

25 ## 2.1 The IAGOS European Research Infrastructure

Since 1994, IAGOS in-situ observations of essential climate variables (temperature, water vapour, and ozone) in the UTLS are provided on a global scale by the European Research Infrastructure IAGOS (Petzold et al., 2015). IAGOS builds on the former EU framework projects MOZAIC and CARIBIC (Civil Aircraft for the Regular Investigation of the Atmosphere Based on an Instrument Container, Brenninkmeijer et al., 2007).

30

Currently (2017) up to ten passenger aircraft from various international airlines are equipped with scientific instruments to monitor the meteorological state (temperature, water vapor and wind) and atmospheric chemical composition. Besides the measurements of temperature and (relative) humidity, the IAGOS-CORE Ppackage 1 includes instruments to measure ozone and carbon monoxide. Ozone is measured by UV absorption (Thermo Scientific for Model 49), and CO by IR-absorption (Thermo Scientific for Model 48 Trace Level). Both instruments are regularly calibrated before and after deployment and the overall uncertainty for ozone is 2 ppb ±2% and for CO 5 ppb ±5% (Thouret et al., 1998; Nédélec et al., 2015). Additionally, positions, pressure, ambient temperature (measured from the aircraft), aircraft speed, wind speed and wind direction are provided by the A330 and A340 aircraft avionic system, respectively (details are given by Petzold et al. (2015)). For completeness, several additionally parameters (NO_y, NO, NO_2) are measured or will be soon measured (CH_4, CO_2) from IAGOS Packages 2 which are described by Petzold et al. (2015).

2.2 Temperature measurements and evaluations

On IAGOS aircraft, temperature is measured in-situ with a compact airborne sensing device AD-FS2 (Aerodata, Braunschweig Germany), which is installed in an appropriate aeronautic housing (Helten et al., 1998). From 1994 to 2009 a platinum resistance sensor (Pt100) was attached at the humidity sensing device. After a design change from MOZAIC to IAGOS this sensor is now directly at the humidity sensor (Figure 1Figure 1). Inter-comparison between both systems showed a temperature deviation of less than 0.1 K in the calibration chamber, which is below the sensor uncertainty. The temperature is measured with an uncertainty of ±0.25 K by a microprocessor controlled transmitter unit (Model HMT 333, Vaisala, Finland), which passes the signal to the data acquisition system of IAGOS Package 1 instrument, where it is recorded with a time resolution of 4 s (Nédélec et al., 2015; Petzold et al., 2015). Pre- and post-deployment calibrations from the laboratory are used to evaluate the temperature signal and to ensure the quality of the temperature measurement (Helten et al. 1998, Neis et al. 2015). Typical deployment phases are in the range of 2 to 3 months. Accounting for corrections of adiabatic compression at the inlet part of the housing (Stickney et al., 1994; Moninger et al., 2003), the overall uncertainty of the ambient air temperature is ±0.5 K. More details are given in the standard operation procedure (SOP) of the IAGOS capacitive Hygrometer (ICH), available at www.iagos.org. For the purpose of this analysis the dataset was reduced to 1-minute averages.

To ensure the reliability of the IAGOS temperature observations, we make use of temperature measurements from the AIRTOSS-ICE aircraft campaign (Aircraft Towed Sensor Shuttle – Inhomogeneous Cirrus Experiment) which focused on mid-latitude cirrus clouds (Neis et al., 2015). During this campaign, the IAGOS temperature instrument was installed on the research aircraft (Learjet 35A), and provides the opportunity to compare both temperature measurements. The aircraft temperature measurement was made with Pt-100 thermistor mounted in the same type of Rosemount as for the IAGOS temperature measurements. The temperature sensor of the research aircraft has been calibrated regularly with an uncertainty of about 0.5 to 1.0 K. In general, the research aircraft flies only a certain amount of hours, and we expect that the

temperature measurement of the research aircraft are regular quality controlled and are well calibrated. Figure 2Figure 2 shows the temperature correlation and the temperature bias (ΔT) at pressures below 400 hPa during seven flights. The general behavior between both temperature measurements agreed well along the flight tracks. The mean deviation is $\Delta T = -0.3$ K, with a pressure dependency to a smaller temperature bias towards 200 hPa. The temperature correlation is high and the temperature bias is smaller than the overall uncertainty of 0.5 K (sensor and adiabatic compression correction), which demonstrates the capability of the IAGOS temperature sensor to measure the ambient temperature at cruise altitude very precisely.

The temperature measured from the aircraft (T_{AC}) is based on total air temperature (TAT) designed for subsonic aircraft (Goodrich Corporation, formerly Rosemount Aerospace). The total air temperature is defined as the ambient air temperature plus the temperature increase due to adiabatic compression in the Rosemount housing. Typically three TAT-sensors (platinum resistance sensor) are installed on at the nose region of the aircraft, but in general only one is used for the pilots and stored for IAGOS. The other two sensors are used to monitor the differences between all TAT-sensors. In general, the airline follow the AMDAR quality recommendations (WMO, 2003), and the TAT-sensors are regular checked by visible inspection. An exchange of one TAT-sensor is performedmaintained, if it differs more than 3°C from the other two TAT-sensors. For the Lufthansa fleet, T_{AC} is compared additionally to models from the German Weather Service to ensure an accuracy of ±3°C (pers. Communication, Lufthansa Technik AG).

2.3 Spatial and temporal data coverage

In this study all temperature measurements from the IAGOS-CORE flights at cruise altitude (p <350 hPa) from January 1995 to December 2012 are used. More recent measurements are not yet validated and therefore not included in this study, therefore most of the IAGOS temperature observations rely on the old former instrument design. Following previous IAGOS or MOZAIC analyses (Thouret et al., 2006; Dyroff et al., 2015; Thomas et al., 2015; Stratmann et al., 2016), we divide the data into 14 geographical regions (Figure 3Fig. 3 and Table 1Tab. 1). Seasonal and regional differences of the temperature behavior can then be linked to the different dynamical patterns. The largest amount of measurements are obtained in the mid-latitudes (North America, North Atlantic, Europe and Central Asia), whereas the amount of measurements in the northern latitudes (North Canada, Greenland, Scandinavia, North Asia) and the tropical regions (Middle America, Tropical Atlantic, North Africa, Tropical Asia, South America, South Africa) is much smaller and does not provide a continuous coverage over the presented period. The data coverage for each region over the analyzed period is shown in Figure 4Figure 4.

Changes of radiatively active chemical constituents (e.g. ozone and water vapor) within the LMS lowermost stratosphere (LMS) and the upper troposphere (UT) have different impacts on the ambient temperature in these layers. In general the tropopause layer (TPL) separates the stable stratified LMS and the unstable UT. In the present study, the pressure of the thermal tropopause ($P_{TPHWMO}p_{TPHWMO}$) is derived from the ERA-Interim (see below), and it is used to determine the position

of the aircraft relative to this layer, and to distinguish if the aircraft flew within the UT, in the TPL or the LMS. This is achieved using the following criteria:

LMS : $p < p_{TPHWMO}$ - 15 hPa, which is limited by the maximum cruise altitude (p ~ 190 hPa)

5   TPL: $p = p_{TPHWMO} \pm 15$ hPa

UT: $p > p_{TPHWMO} + 15$ hPa, limited to 350 hPa

~~LMS : p < max(TPL), which is limited by the maximum cruise altitude (p ~ 190 hPa)~~

~~TPL: p = P_TPHWMO ± 15 hPa~~

~~UT: p > min(TPL), limited to 350 hPa~~

10

A comparable definition has been used by Thouret et al. (2006), where the pressure of the 2-pvu-surface was used to define the dynamical tropopause layer with a vertical depth of ±15 hPa. ~~The tropopause height can be estimated by several other approaches and is then called: the dynamical tropopause (Holten, 2004), the chemical tropopausePan et al., 2004; Bethan et al., 1996; Sprung and Zahn, 2010) or estimated by a combination of them (Stohl et al., 2003; Cui et al., 2009; Kunz et al.,~~

15 ~~2011; Wolcox et al., 2012). However, depending on the region studied, different thresholds for different approaches are used.~~ The thermal tropopause is valid within all latitude bands, whereas the dynamical tropopause cannot be used in tropical regions, because the Coriolis parameter is zero, therefore potential vorticity (PV) goes to zero and the 2-pvu surface is not defined (Boothe and Homeyer, 2017). ~~The thermal tropopause is valid within all latitude bands, whereas the dynamical tropopause cannot be used in tropical regions, because the Coriolis parameter is not defined near the equator, which is~~

20 ~~needed for the PV calculation (Boothe and Homeyer, 2017).~~

Within the IAGOS project, measurements of ozone (since 1994) and carbon monoxide (since 2002) are available and are used to justify our layer classification scheme. Figure 5~~Figure 5~~ shows the cumulative distribution of ozone, carbon monoxide and potential vorticity (PV). All distributions demonstrate a clear separation between the three layers based on the

25 thermal tropopause. Median ozone mixing rations in the UT are around 60 ppb, median CO mixing ratios are 90 ppb, and median PV is 0.5 PVU, respectively. Within the LMS, the median values for O_3 (310 ppb), for CO (40 ppb) and for PV (7 PVU) are consistent to previous studies (Pan et al., 2004; Kunz et al., 2008; Brioude et al., 2009; Schmidt et al., 2010).

2.4 Meteorological reanalysis

30 ERA-Interim covers the period from 1979 until present, assimilating observational data from various satellites, radiosondes, buoys, commercial aircraft and others (Dee et al., 2011; Simmons et al., 2014). Note that the IAGOS temperature observations are not assimilated in any numerical weather prediction model or reanalysis product, which makes it unique for

model evaluation. For this study, the 6-hourly output from ERA-I (0.75° x 0.75°) were interpolated onto a 1° x 1° horizontal grid, and on 60 vertical levels of constant pressure and potential temperature (Kunz et al., 2014). Additionally, the variables of the potential vorticity (PV), and the pressure of the thermal tropopause (p_{TPHWMO}) based on the WMO criteria were calculated (WMO, 1957; Reichler et al., 2003). The ERA-I data were linearly interpolated (longitude, latitude, pressure, time) onto each flight track with 4 s resolution as described by Kunz et al. (2014). As for IAGOS temperature observations the ERA-I dataset was reduced to 1-minute averages.

3 Results

3.1 IAGOS temperature measurements over different regions

Approximately 69% of the IAGOS temperature measurements in the UTLS were obtained in the mid-latitude band between 30° to 60°N. Therefore, we show our detailed analysis within all three layers of the UTLS only over the North Atlantic region between January 1995 and December 2012 and provide additional material as supplement. Figure 6Figure 6 shows monthly median temperature of the IAGOS observations in the LMS, TPL and UT for the period from 1995 to 2012 over the North Atlantic region. All layers show a seasonal temperature variation of 5 to 10 K in each year. The warmest temperatures are observed within the LMS and the coldest (as expected) within the TPL. Between 2006 and 2008, the wintertime temperature minima at the TPL and the UT are almost 3 to 4 K warmer compared to the other years. These enhanced temperatures can also be observed over Europe and Central Asia within these layers. All other regions in the extratropics show mostly comparable time series in amplitude and seasonal variations to the North Atlantic region, if enough measurements were available for each vertical layer. Throughout all regions in the extratropics, the summer / winter months (JJA / DJF) are always the warmest / coldest within all three layers (see tables in the supplement). Over the tropical regions, no measurements are available in the LMS and TPL, because the TPL is mostly above cruise altitude. In the UT, the temperature is mostly constant throughout the year for each tropical region respectively.

3.2 Temperature anomalies and trends

The temperature trends are derived from using a robust regression analysis over the full period of deseasonalized temperature observations. Eighteen-year monthly averages (e.g. mean of all January values, mean of all February values, etc.) were subtracted from the time series of each layer in all regions. The temperature trends exhibit no significant non-linear contributions. We ensured that the data is homoscedastic, and it is nearly normally distributed (median and mean values are very close, see table in supplements). Additional, we checked that the data is not auto-correlated. Temperature trends are reported only if at least 90% of all months have at least 200 data points each (Figure 4Figure 4). The Mann-Kendall-test is used to identify the trend significance (Mann, 1945; Kendall, 1975; Gilbert, 1987). The 90% threshold was chosen because the Mann-Kendall significance of the trend analysis did not change, when 10% of the data were randomly

excluded from the trend calculation. The robustness of temperature trends was tested by skipping the first or last year of the 18-year period. Within all layers and all regions each trend keeps the same sign and the trend values varied within the standard error. The only exception was the upper troposphere over North America where the temperature trend changed from slight positive trend (18 years) to neutral when the final year was removed.

5

Figure 7Figure 7 shows the time series of temperature anomalies and linear trend lines over the North Atlantic region using the IAGOS observations. The monthly temperature anomalies in each layer vary by ±3 K (strongest within the UT). Temperature trends are +0.22 (±0.20) K/dec in the UT, +0.25 (±0.16) K/dec in the TPL and -0.05 (±0.17) K/dec in the LMS. None of the trends are significant at the 95% level. In the UT, the amplitudes of the

10 smoothed time series are larger and decrease towards the upper levels. They show two warmer phases and two colder phases during the analyzed period.

Figure 8Figure 8 and Table 3Table 3 summarize the temperature trends for different regions in the different layers. Within the LMS, significant cooling is observed over Greenland (-1.39 (±0.29) K/dec) and North America (-0.71 (±0.21) K/dec).

15 The smaller trend over Europe (-0.53 (±0.20) K/dec) is not significant. The other regions don't have enough data for a meaningful trend analysis in the LMS. Within the TPL, there are only three regions with sufficient data coverage, and no significant trend is detected, although there is a weak indication for a small warming over North America and the North Atlantic and for a small cooling over Europe. In the UT, significant cooling is found over North America (-1.08 (±0.18) K/dec), while temperatures over Europe (-0.59 (±0.1) K/dec) decrease not significant. Over the North Atlantic (+0.22

20 (±0.20) K/dec) and over Central Asia (+0.32 (±0.33) K/dec), temperatures increase non-significantly, respectively. Over Tropical Asia (-0.54 (±0.04) K/dec) temperature show a tendency to decrease, but without significance at the 95% level. This result is somehow puzzling, because it is assumed that the temperature increases in the tropics in the UT, equivalent to the surface temperature (Khandu et al., 2016). One reason could be related to the tropopause definition, but O_3 (40 to 80 ppb) and PV (always below 1 PVU) indicate that the selected air mass is tropospheric. Another reason could be related to a higher

25 variability of the temperature due to large scale influence or simply that the data coverage in this region is still too poor in this region, which might lead to a higher variability of the local temperatures which then mask the temperature trend.

4 Discussion

4.1 Inter-comparison with ERA-Interim reanalysis

ERA-interim reanalysis is widely used for intercomparions and is currently the latest global atmospheric reanalysis product

30 provided by ECMWF (Dee et al., 2011; Fujiwara et al., 2017). Simmons et al. (2014) indicated that there is a problem with

the temperature near the ~~local~~ tropopause in the tropics and extratropics. Here, we demonstrate the use of IAGOS observations to evaluate the ERA-I reanalysis within the UTLS region.

Figure 9~~Figure 9~~ shows the inter-comparison between the temperature of the IAGOS observations and the ERA-I reanalysis of all 1-minute mean data over the North Atlantic region at cruise altitude (below 350 hPa). The good agreement between both temperatures is reflected by the high correlation coefficient ($R^2 = 0.97$) and the slope of 0.94 from the regression fit. The bias between the temperature of the IAGOS measurements and ERA-I is -0.02 K over the entire period. The absolute bias is 0.81 ($\pm$0.72) K and 93% of the data is close to the 1:1-line ($\pm$2K). The spread of the correlation corresponds to a larger variability of the IAGOS temperature measurements when the aircraft flew through clouds, or from the interpolation of the gridded ERA-I temperature to the aircraft position. Since 2011, IAGOS measurements with a backscatter cloud probe are available and these measurements will be used in the future to distinguish between clear sky and in-cloud measurements to further reduce the uncertainties. The good agreement between the 1-minute mean temperature data from ERA-I and the IAGOS observation in the other regions is documented in Table 2~~Table 2~~, ~~where~~ when R^2 is always larger than 0.97, the slope of each regression fit is between 0.94 and 0.98, and the absolute bias varies between 0.60 to 0.98 K with a standard deviation of 0.56 to 0.79 K.

4.2 Anomalies and trends using ERA-I

The ERA-I temperature time series and temperature anomalies agree with the IAGOS measurements mostly in phase and amplitude within all vertical layers over the North Atlantic region (Figure 6~~Fig. 6~~ and Figure 7~~Fig. 7~~). Larger deviations can be found at the beginning of the time series in the mid-1990s, where the IAGOS temperatures are warmer than ERA-I, and after 2008, when the opposite occurs. The temperature trends from ERA-I show a slight warming over the North Atlantic region (+0.38 ($\pm$0.18) K/dec) in the UT, and 0.46 ($\pm$0.15) K/dec in the TPL. Neither the trends in the UT nor the trend in the TPL are significant.

In the LMS, ERA-I temperatures show a significant increase of +0.56 ($\pm$0.17) K/dec over the 18 years of the study period at the 95% significance level, which is not the case for the IAGOS observations, where the temperature shows a not significant decrease of -0.05 ($\pm$0.17) K/dec over the analyzed period (see section 4.1 above). We calculated the temperature trends of the lower (5-percentile, colder temperatures) and upper (95-percentile warmer temperature) range of the data to determine the robustness of these trends within the LMS. For the 5-percentiles, the temperature trend from ERA-I remains significant (99% confidence) at +0.48 K/dec, while the trend from the IAGOS observations is still not significantly decreasing at -0.22 K/dec (87% confidence). In the upper range, the differences between the temperature trends diminish to +0.12 K/dec (53% confidence) using ERA-I and -0.07 K/dec (73% confidence) using the IAGOS observations. This shows that within the colder temperature regimes, the trend difference is significant, while for the warmest temperatures regimes, the variability is too large to obtain a significant temperature trend.

9

Figure 8Figure 8 and Table 3Table 3 also contain the temperature trend estimates for ERA-Interim. In contrast to the IAGOS observations, a cooling tendency in the LMS can only be found over Greenland of (-0.79 (±0.29) K/dec), and over North America (-0.25 (±0.21) K/dec), but both are not significant. The strong cooling trend over Greenland results from elevated temperatures in the late 1990s. Over Europe, where IAGOS observations showed an not insignificant decrease, ERA-I exhibits a very small (and not significant) warming of +0.11 (±0.19) K/dec.

Within the TPL, the temperatures trends are mostly comparable to the IAGOS observations (expect over Europe), but all trends are not significant. The best consistency between the temperature trends from IAGOS observations and ERA-I is seen for the UT layer. Except over Europe, where ERA-I shows slightly less cooling thean IAGOS observations, and Central Asia, where ERA-I has a stronger warming trend.

4.3 Possible sources of the differing temperature trends

There is a remarkable coherence of the temperature time series and the anomalies of the observations and ERA-I in all regions and layers, but the linear temperature trends reveal a deviation between the two data sets, especially in the LMS. In order to investigate this in more detail, Figure 10Figure 10 shows the annual averages of the monthly deviation between the IAGOS observations and ERA-I at the LMS over eight regions. From 1995 to 2002 the observations are always warmer than ERA-I within the Northern Hemisphere. The sign changes to neutral until 2008, and since 2009 the IAGOS observations are colder than ERA-I. This temporally varying temperature bias is one reason why the temperature trends are not equal between the two data sets. The differences cannot be explained by the chosen definition to determine the different layers within the UTLS region. For example, ozone observations show seasonal variations with values between 350 to 640 ppb, which are robust values for stratospheric air masses with a typical seasonality (Brioude et al., 2009). Additionally we use CO observations (from 2002 to 2008) and PV with similar results.

Furthermore, we demonstrated that temperature measurements from IAGOS agree well to quality controlled measurements aboard research aircraft (Figure 2Figure 2). Therefore, we expect that the trends from IAGOS are robust, which leads to the hypothesis that ERA-I exhibits temperature biases which vary with time.

Some of the observed deviations can be explained by changes of the data assimilation sources in ERA-I, which employed several new satellite products at various times after the year 2000. Simmons et al. (2014) showed that the source of input data (e.g. inconsistent SST bias) changed in ERA-I after 2002. Since 2001, temperature profiles were assimilated from GPSRO measurements into ERA-I, which lead to a cold bias in ERA-I, where the largest effects appeared after 2006 in ERA-I, when the amount of assimilated data increased. This could lead to a warming of the lower stratosphere and a cooling at around 200 hPa (Poli et al. 2008; Poli et al. 2010).

Another source of the deviation could be expected from the increasing number of temperature measurements from the AMDAR system onboard of passenger aircraft. Ballish and Kumar (2008) showed that the temperatures from passenger aircraft are warm-biased at cruise altitude (200 to 300 hPa). This has not been accounted for ERA-I. Furthermore, the WMO reported that the amount of aircraft reports increased from 100 000 in early 2000 to more than 350 000 after 2012, and became the 3rd most important data source of assimilation for short term forecast in numerical weather prediction (WMO, 2014; Petersen et al., 20156; WMO, 2014).

In order to test the second hypothesis, Figure 7Figure 7 also includes the temperature trend derived over the North Atlantic region using the aircraft temperature measurements (T$_{AC}$), which we assume to be comparable to AMDAR measurements. The temperature trend in the LMS is also positive, and in-between the neutral temperature trend using the IAGOS temperature observations and the positive temperature trend using ERA-I. This gives an indication that the assimilated aircraft measurements could indeed be a cause for bias in the ERA-I data. It is therefore planned to introduce a bias correction for AMDAR aircraft observations in the next ERA reanalysis (ERA5, personal communication D. Dee).

5 Conclusion

In this study, nearly two decades of in-situ IAGOS temperature measurements using passenger aircraft by the European Research Infrastructure IAGOS were presented and used to determine regional trends in the UTLS. The quality of the temperature measurements is regularly evaluated in the laboratory through pre- and post-deployment calibration, and it has been assessed by intercomparison with temperature observations from research aircraft.

UTLS temperature time series and trends are analyzed for 14 different regions within the northern latitudes, mid-latitudes and tropics, and separated into the lowermost stratosphere, the tropopause layer and the upper troposphere using the thermal tropopause from ERA-I as reference layer.

The inter-annual variability within all regions and layers is mostly consistent between the IAGOS observations and the ERA-I during the past 18 years (1995 to 2012), which is not the case for the temperature trends. For the temperature trend only regions are considered when at least 90% of all months are available. Over the North Atlantic where the largest amount of measurements is available, we found a significant (95% confidence level) positive trend (+0.56 K/dec) in ERA-I and a small negative trend (-0.05 K/dec) in the observations (not significant) in the LMS. A significant (95% confidence) negative temperature trend over Greenland and North America in the LMS for the IAGOS temperature is found. Over Europe both temperatures trends are not significant, but the temperature trend from the IAGOS observations is negative and for ERA-I positive. Within the tropopause layer we found mostly comparable trends, expect over Europe, where the sign of trends is different, but all trends are not significant. The calculated temperature trends in the UT show all the same sign for each

region, whereas only a significant trend is found over North America with a large cooling rate of about -0.92 K/dec (ERA-I) and -1.08 K/dec (IAGOS) in the last 18 years.

The large deviation between the different LMS temperature trends from IAGOS observations and ERA-I data is mostly related to the temporally varying bias between both temperature time series. As we have no reason to assume an evolving bias in the IAGOS observations, we conclude that ERA-I temperatures are too cold between 1995 and 2001, and too warm after 2007. These dates roughly correspond with changes in the data streams that were used in the ERA-I data assimilation. The evolution of ERA-I temperature can be explained by additional assimilation of GPSRO data and data from the AMDAR passenger aircraft network, which have been shown to be warm-biased. These data sources play an increasing role in the ERA-I data assimilation after 2006. Our recommendation is therefore to include a bias correction for these temperature measurements in future versions of the European reanalysis and use IAGOS observations as anchor point.

The IAGOS temperature measurements highlight the need of independent global measurements with a high and long-term accuracy to quantify long-term changes especially in the UTLS region, and help to identify inconsistencies between different datasets of observations and models. Due to the expansion of the IAGOS aircraft fleet with airlines from other continents, we are looking forward to investigate temperature changes in the UTLS over several more regions in a few years from now.

Author contributions

F.B. conducted the analysis of the measurements and wrote the manuscript. H.G.J.S., P.N., S.R., and U.B. were in charge of the instrument setup, calibration and processing of the measurements. A.P., M.G.S. and A.W. helped with the interpretation of the data and in the manuscript writing process. P.N., D.B. and V.T. were in charge to provide the CO and ozone measurements of MOZAIC and IAGOS.

Competing interests

The authors declare that they have no conflict of interests. The measurements are available free of charge on www.iagos.org.

References

Ballish, B. A. and Kumar, V. K.: Systematic differences in aircraft and radiosonde temperatures, Bull. Am. Meteorol. Soc., 89(11), 1689–1707, doi:10.1175/2008BAMS2332.1, 2008.

Bencherif, H., Diab, R. D., Portafaix, T., Morel, B., Keckhut, P. and Moorgawa, A.: Temperature climatology and trend estimates in the UTLS region as observed over a southern subtropical site, Durban, South Africa, Atmos. Chem. Phys., 6, 5121–5128, doi:10.5194/acpd-6-1301-2006, 2006.

Blunden, J., Arndt, D. S. and Espinoza, J. C.: State of the climate in 2013, Bull. Am. Meteorol. Soc., 95(7), doi:10.1175/2014BAMSStateoftheClimate.1, 2014.

Boothe, A. C. and Homeyer, C. R.: Global large-scale stratosphere–troposphere exchange in modern reanalyses, Atmos. Chem. Phys., 17(9), 5537–5559, doi:10.5194/acp-17-5537-2017, 2017.

Bortz, S. E., Prather, M. J., Cammas, J.-P., Thouret, V. and Smit, H.: Ozone, water vapor, and temperature in the upper tropical troposphere: Variations over a decade of MOZAIC measurements, J. Geophys. Res., 111(D5), D05305, doi:10.1029/2005JD006512, 2006.

Brenninkmeijer, C. A. M., Crutzen, P., Boumard, F., Dauer, T., Dix, B., Ebinghaus, R., Filippi, D., Fischer, H., Franke, H., Frieβ, U., Heintzenberg, J., Helleis, F., Hermann, M., Kock, H. H., Koeppel, C., Lelieveld, J., Leuenberger, M., Martinsson, B. G., Miemczyk, S., Moret, H. P., Nguyen, H. N., Nyfeler, P., Oram, D., O'Sullivan, D., Penkett, S., Platt, U., Pupek, M., Ramonet, M., Randa, B., Reichelt, M., Rhee, T. S., Rohwer, J., Rosenfeld, K., Scharffe, D., Schlager, H., Schumann, U., Slemr, F., Sprung, D., Stock, P., Thaler, R., Valentino, F., van Velthoven, P., Waibel, A., Wandel, A., Waschitschek, K., Wiedensohler, A., Xueref-Remy, I., Zahn, A., Zech, U. and Ziereis, H.: Civil Aircraft for the regular investigation of the atmosphere based on an instrumented container: The new CARIBIC system, Atmos. Chem. Phys., 7(18), 4953–4976, doi:10.5194/acp-7-4953-2007, 2007.

Brioude, J., Cammas, J. P., Cooper, O. R. and Nedelec, P.: Characterization of the composition, structure, and seasonal variation of the mixing layer above the extratropical tropopause as revealed by MOZAIC measurements, J. Geophys. Res. Atmos., 114(7), 1–17, doi:10.1029/2007JD009184, 2009.

Dee, D. P., Uppala, S. M., Simmons, a. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. a., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, a. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, a. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137(656), 553–597, doi:10.1002/qj.828, 2011.

Drüe, C., Frey, W., Hoff, A. and Hauf, T.: Aircraft type-specific errors in AMDAR weather reports from commercial aircraft, Q. J. R. Meteorol. Soc., 134(630), 229–239, doi:10.1002/qj.205, 2008.

Dyroff, C., Zahn, A., Christner, E., Forbes, R., Tompkins, A. M. and van Velthoven, P. F. J.: Comparison of ECMWF analysis and forecast humidity data with CARIBIC upper troposphere and lower stratosphere observations, Q. J. R. Meteorol. Soc., 141(688), 833–844, doi:10.1002/qj.2400, 2015.

Fueglistaler, S., Abalos, M., Flannaghan, T. J., Lin, P. and Rande, W. J.: Variability and trends in dynamical forcing of tropical lower stratospheric temperatures, Atmos. Chem. Phys., 14(24), 13439–13453, doi:10.5194/acp-14-13439-2014, 2014.

Fujiwara, M., Wright, J. S., Manney, G. L., Gray, L. J., Anstey, J., Birner, T., Davis, S., Gerber, E. P., Lynn Harvey, V., Hegglin, M. I., Homeyer, C. R., Knox, J. A., Krüger, K., Lambert, A., Long, C. S., Martineau, P., Molod, A., Monge-Sanz, B. M., Santee, M. L., Tegtmeier, S., Chabrillat, S., Tan, D. G. H., Jackson, D. R., Polavarapu, S., Compo, G. P., Dragani, R., Ebisuzaki, W., Harada, Y., Kobayashi, C., McCarty, W., Onogi, K., Pawson, S., Simmons, A., Wargan, K., Whitaker, J. S. and Zou, C. Z.: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems, Atmos. Chem. Phys., 17(2), 1417–1452, doi:10.5194/acp-17-1417-2017, 2017.

Gettelman, A., Pan, L. L., Randel, W. J., Hoor, P., Birner, T. and Hegglin, M. I.: the Extratropical Upper Troposphere and Lower Stratosphere, Rev. Geophys., 49(3), 1–31, doi:10.1029/2011RG000355.1.INTRODUCTION, 2011.

Gilbert, R. O.: Statistical Methods for Environmental Pollution Monitoring, Wiley, New York., 1987.

Heise, S., Wickert, J., Beyerle, G., Schmidt, T., Smit, H., Cammas, J.-P. and Rothacher, M.: Comparison of Water Vapor and Temperature Results From GPS Radio Occultation Aboard CHAMP With MOZAIC Aircraft Measurements, IEEE Trans. Geosci. Remote Sens., 46(11), 3406–3411, doi:10.1109/TGRS.2008.920268, 2008.

Helten, M., Smit, H. G. J., Sträter, W., Kley, D., Nedelec, P., Zöger, M. and Busen, R.: Calibration and performance of automatic compact instrumentation for the measurement of relative humidity from passenger aircraft, J. Geophys. Res., 103(D19), 25643, doi:10.1029/98JD00536, 1998.

Ho, S.-P., Peng, L. and Vömel, H.: Characterization of the long-term radiosonde temperature biases in the upper troposphere and lower stratosphere using COSMIC and Metop-A/GRAS data from 2006 to 2014, Atmos. Chem. Phys., 17(7), 4493–4511, doi:10.5194/acp-17-4493-2017, 2017.

Ingleby, B., Isaksen, L. and Dahoui, M.: ECMWF steps up assimilation of aircraft weather data. [online] Available from: http://www.ecmwf.int/en/newsletter/148/news/ecmwf-steps-assimilation-aircraft-weather-data, 2016.

Kendall, M. G.: Rank Correlation Methods, 4th edition., Griffin, Charles, London., 1975.

Khandu, Awange, J. L. and Forootan, E.: Interannual variability of temperature in the UTLS region over Ganges–Brahmaputra–Meghna river basin based on COSMIC GNSS RO data, Atmos. Meas. Tech., 9(4), 1685–1699, doi:10.5194/amt-9-1685-2016, 2016.

Kunz, A., Schiller, C., Rohrer, F., Smit, H. G. J., Nedelec, P. and Spelten, N.: Statistical analysis of water vapour and ozone in the UT/LS observed during SPURT and MOZAIC, Atmos. Chem. Phys., 8(22), 6603–6615, doi:10.5194/acp-8-6603-

14

2008, 2008.

Kunz, A., Spelten, N., Konopka, P., Müller, R., Forbes, R. M. and Wernli, H.: Comparison of Fast In situ Stratospheric Hygrometer (FISH) measurements of water vapor in the upper troposphere and lower stratosphere (UTLS) with ECMWF (re)analysis data, Atmos. Chem. Phys., 14(19), 10803–10822, doi:10.5194/acp-14-10803-2014, 2014.

5   ~~Kuo, Y. H., Sokolovskiy, S. V, Anthes, R. a and Vandenberghe, F.: Assimilation of GPS radio occultation data for numerical weather prediction, Terr. Atmos. Ocean. Sci., 11(1), 157–186, 2000.~~

Kursinski, E. R., Hajj, G. A., Schofield, J. T., Linfield, R. P. and Hardy, K. R.: Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System, J. Geophys. Res. Atmos., 102(D19), 23429–23465, doi:10.1029/97JD01569, 1997.

10   Mann, H. B.: Nonparametric tests against trend, Econometrica, 13, 163–171, 1945.

Marenco, A., Thouret, V., Nédélec, P., Smit, H., Helten, M., Kley, D., Karcher, F., Simon, P., Law, K., Pyle, J., Poschmann, G., Von Wrede, R., Hume, C. and Cook, T.: Measurement of ozone and water vapor by Airbus in-service aircraft: The MOZAIC airborne program, an overview, J. Geophys. Res. Atmos., 103(D19), 25631–25642, doi:10.1029/98JD00977, 1998.

15   Moninger, W. R., Mamrosh, R. D. and Pauley, P. M.: Automated meteorological reports from commercial aircraft, Bull. Am. Meteorol. Soc., 84(2), 203–216+149, doi:10.1175/BAMS-84-2-203, 2003.

Nédélec, P., Blot, R., Boulanger, D., Athier, G., Cousin, J.-M., Gautron, B., Petzold, A., Volz-Thomas, A. and Thouret, V.: Instrumentation on commercial aircraft for monitoring the atmospheric composition on a global scale: the IAGOS system, technical overview of ozone and carbon monoxide measurements, Tellus B, 67(MOZAIC/IAGOS), 1–16, 20   doi:10.3402/tellusb.v67.27791, 2015.

Neis, P., Smit, H. G. J., Rohs, S., Bundke, U., Krämer, M., Spelten, N., Ebert, V., Buchholz, B., Thomas, K. and Petzold, A.: Quality assessment of MOZAIC and IAGOS capacitive hygrometers: insights from airborne field studies, Tellus B, 67, 1–11, doi:10.3402/tellusb.v67.28320, 2015.

Pan, L. L., Randel, W. J., Gary, B. L., Mahoney, M. J. and Hintsa, E. J.: Definitions and sharpness of the extratropical 25   tropopause: A trace gas perspective, J. Geophys. Res. D Atmos., 109(23), 1–11, doi:10.1029/2004JD004982, 2004.

Petersen, R. A.: On the Impact and Benefits of AMDAR Observations in Operational Forecasting—Part I: A Review of the Impact of Automated Aircraft Wind and Temperature Reports, Bull. Am. Meteorol. Soc., 97(4), 585–602, doi:10.1175/BAMS-D-14-00055.1, 2016.

Petzold, A., Thouret, V., Gerbig, C., Zahn, A., Brenninkmeijer, C. A. M., Gallagher, M., Hermann, M., Pontaud, M., Ziereis, 30   H., Boulanger, D., Marshall, J., Nédélec, P., Smit, H. G. J., Friess, U., Flaud, J.-M., Wahner, A., Cammas, J.-P. and Volz-Thomas, A.: Global-scale atmosphere monitoring by in-service aircraft – current achievements and future prospects of the European Research Infrastructure IAGOS, Tellus B, 67, 1–24, doi:10.3402/tellusb.v67.28452, 2015.

Poli, P., Healy, S. B., Rabier, F. and Pailleux, J.: Preliminary assessment of the scalability of GPS radio occultations impact in numerical weather prediction, Geophys. Res. Lett., 35(23), L23811, doi:10.1029/2008GL035873, 2008.

Poli, P., Healy, S. B. and Dee, D. P.: Assimilation of Global Positioning System radio occultation data in the ECMWF ERA-Interim reanalysis, Q. J. R. Meteorol. Soc., 136(653), 1972–1990, doi:10.1002/qj.722, 2010.

Randel, W. J., Shine, K. P., Austin, J., Barnett, J., Claud, C., Gillett, N. P., Keckhut, P., Langematz, U., Lin, R., Long, C., Mears, C., Miller, A., Nash, J., Seidel, D. J., Thompson, D. W. J., Wu, F. and Yoden, S.: An update of observed stratospheric temperature trends, J. Geophys. Res., 114(D2), D02107, doi:10.1029/2008JD010421, 2009.

Reichler, T., Dameris, M. and Sausen, R.: Determining the tropopause height from gridded data, Geophys. Res. Lett., 30(20), 1–5, doi:10.1029/2003GL018240, 2003.

Rieckh, T., Scherllin-Pirscher, B., Ladstädter, F. and Foelsche, U.: Characteristics of tropopause parameters as observed with GPS radio occultation, Atmos. Meas. Tech., 7(11), 3947–3958, doi:10.5194/amt-7-3947-2014, 2014.

Riese, M., Ploeger, F., Rap, A., Vogel, B., Konopka, P., Dameris, M. and Forster, P.: Impact of uncertainties in atmospheric mixing on simulated UTLS composition and related radiative effects, J. Geophys. Res. Atmos., 117(D16), n/a-n/a, doi:10.1029/2012JD017751, 2012.

Santer, B. D., Mears, C., Doutriaux, C., Caldwell, P., Gleckler, P. J., Wigley, T. M. L., Solomon, S., Gillett, N. P., Ivanova, D., Karl, T. R., Lanzante, J. R., Meehl, G. A., Stott, P. A., Taylor, K. E., Thorne, P. W., Wehner, M. F. and Wentz, F. J.: Separating signal and noise in atmospheric temperature changes : The importance of timescale, J. Geophys. Res., 116, 1–19, doi:10.1029/2011JD016263, 2011.

Schmidt, T., Cammas, J.-P., Smit, H. G. J., Heise, S., Wickert, J. and Haser, A.: Observational characteristics of the tropopause inversion layer derived from CHAMP/GRACE radio occultations and MOZAIC aircraft data, J. Geophys. Res., 115(D24), D24304, doi:10.1029/2010JD014284, 2010.

Seidel, D. J. and Randel, W. J.: Variability and trends in the global tropopause estimated from radiosonde data, J. Geophys. Res., 111(D21), D21101, doi:10.1029/2006JD007363, 2006.

Seidel, D. J., Li, J., Mears, C., Moradi, I., Nash, J., Randel, W. J., Saunders, R., Thompson, D. W. J. and Zou, C.: Stratospheric temperature changes during the satellite era, J. Geophys. Res. Atmos., 121(2), 664–681, doi:10.1002/2015JD024039, 2016.

Simmons, A. J., Poli, P., Dee, D. P., Berrisford, P., Hersbach, H., Kobayashi, S. and Peubey, C.: Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim, Q. J. R. Meteorol. Soc., 140(679), 329–353, doi:10.1002/qj.2317, 2014.

Solomon, S., Rosenlof, K. H., Portmann, R. W., Daniel, J. S., Davis, S. M., Sanford, T. J. and Plattner, G.-K.: Contributions of stratospheric water vapor to decadal changes in the rate of global warming, Science (80-.)., 327(5970), 1219–1223, doi:10.1126/science.1182488, 2010.

Steiner, A. K., Kirchengast, G., Lackner, B. C., Pirscher, B., Borsche, M. and Foelsche, U.: Atmospheric temperature change detection with GPS radio occultation 1995 to 2008, Geophys. Res. Lett., 36(18), L18702, doi:10.1029/2009GL039777, 2009.

Stickney, T. M., Shedlov, M. W. and Thompson, D. I.: Total Temperature Sensors, Rosemount Aerosp. / Goodrich, 1994.

Stratmann, G., Ziereis, H., Stock, P., Brenninkmeijer, C. A. M., Zahn, A., Rauthe-Schöch, A., Velthoven, P. V., Schlager, H.

16

and Volz-Thomas, A.: NO and NOy in the upper troposphere: Nine years of CARIBIC measurements onboard a passenger aircraft, Atmos. Environ., 133, 93–111, doi:10.1016/j.atmosenv.2016.02.035, 2016.

Thomas, K., Berg, M., Boulanger, D., Houben, N., Gressent, A., Nédélec, P., Pätz, H.-W., Thouret, V. and Volz-Thomas, A.: Climatology of NO y in the troposphere and UT/LS from measurements made in MOZAIC, Tellus B, 67, 1–16, doi:10.3402/tellusb.v67.28793, 2015.

Thouret, V., Marenco, A., Logan, J. A., Nedelec, P. and Grouhel, C.: Comparisons of ozone measurements from the MOZAIC airborne program and the ozone sounding network at eight locations, J. Geophys. Res., 103(D19), 25695–25720, doi:10.1029/98JD02243, 1998.

Thouret, V., Cammas, J.-P., Sauvage, B., Athier, G., Zbinden, R., Nédélec, P., Simon, P. and Karcher, F.: Tropopause referenced ozone climatology and inter-annual variability (1994–2003) from the MOZAIC programme, Atmos. Chem. Phys., 6(4), 1033–1051, doi:10.5194/acp-6-1033-2006, 2006.

Wickert, J., Reigber, C., Beyerle, G., König, R., Marquardt, C., Schmidt, T., Grunwaldt, L., Galas, R., Meehan, T. K., Melbourne, W. G. and Hocke, K.: Atmosphere sounding by GPS radio occultation: First results from CHAMP, Geophys. Res. Lett., 28(17), 3263–3266, doi:10.1029/2001GL013117, 2001.

WMO: Definition of the tropopause, WMO Bull., 6, 1957.

WMO: The Benefits of AMDAR Data to Meteorology and Aviation., 2014.

WMO (World Meteorological Organization): Aircraft Meteorological Data Relay (AMDAR) Reference Manual. Technical Report WMO-No. 958., 2003.

Xu, J. and Powell, A. M.: Ensemble spread and its implication for the evaluation of temperature trends from multiple radiosondes and reanalyses products, Geophys. Res. Lett., 37(17), 1–6, doi:10.1029/2010GL044300, 2010.

Table 1: Definition of different regions covered by IAGOS flight tracks.

	Region	Longitude	Latitude
Northern latitudes	North Canada	130°W - 65°W	60°N - 80°N
	Greenland	65°W - 5°W	60°N - 80°N
	Scandinavia	5°W - 45°E	60°N - 80°N
	North Asia	45°E - 180°E	60°N - 80°N
Mid-latitudes	North America	130°W - 65°W	30°N - 60°N
	North Atlantic	65°W - 5°W	30°N - 60°N
	Europe	5°W - 45°E	30°N - 60°N
	Central Asia	45°E - 180°E	30°N - 60°N
Northern Tropics	Middle America	130°W - 65°W	0°N - 30°N
	Tropical Atlantic	65°W - 5°W	0°N - 30°N
	North Africa	5°W - 45°E	0°N - 30°N
	Tropical Asia	45°E - 180°E	0°N - 30°N
Southern Tropics	South America	65°W - 5°W	30°S - 0°N
	South Africa	5°W - 45°E	30°S - 0°N

**Table 2: Statistical parameters summarized for all studied regions (see Figure 9~~Figure 9~~). The slopes and intercepts are determined from the linear robust regression fit between ERA-I and IAGOS temperature. Additionally, for each region the absolute bias and its standard deviation are given.**

	Region	Number of data points $*[10^6]$	R^2	Slope	Intercept	Absolute Bias [K]	Standard Deviation [K]
Northern latitudes	North Canada	0.23	0.98	0.96	9.68	0.73	0.64
	Greenland	0.50	0.97	0.96	9.83	0.81	0.70
	Scandinavia	0.16	0.97	0.96	9.24	0.79	0.68
	North Asia	0.61	0.97	0.95	10.34	0.92	0.77
Mid-latitudes	North America	1.42	0.98	0.95	10.30	0.79	0.69
	North Atlantic	3.23	0.97	0.94	12.26	0.81	0.72
	Europe	2.80	0.97	0.95	10.27	0.85	0.72
	Central Asia	1.76	0.98	0.97	6.35	0.92	0.79
Northern Tropics	Middle America	0.03	0.99	0.97	6.64	0.81	0.57
	Tropical Atlantic	0.39	0.97	0.97	7.29	0.77	0.63
	North Africa	0.71	0.98	0.97	6.53	0.66	0.61
	Tropical Asia	0.50	0.98	0.96	10.49	0.98	0.79
Southern Tropics	South America	0.14	0.98	0.97	6.62	0.87	0.65
	South Africa	0.41	0.99	0.98	4.74	0.60	0.56

5

Table 3: Temperature trends of ERA-I and from the IAGOS observations within the LMS TPL and UT, where at least 90% of measurements were over the entire period (1995 to 2012) available. SE is the standard error of the temperature trend. The temperature trend is significant (Sig=1) if the p-value (consistency) is smaller than 0.06 (>94%), which was derived from the Mann-Kendall-test.

Region	ERA-I				IAGOS			
	ΔT_{18yr} K/dec	SE K/dec	Sig.	p-value	ΔT_{18yr} K/dec	SE K/dec	Sig.	p-value
LMS								
Greenland	-0.79	0.29	0	0.12	-1.39	0.29	1	0.01
North America	-0.25	0.21	0	0.46	-0.71	0.21	1	0.02
North Atlantic	+0.56	0.17	1	0.05	-0.05	0.17	0	0.98
Europe	+0.11	0.19	0	0.83	-0.53	0.20	0	0.17
TPL								
North America	+0.29	0.19	0	0.19	+0.23	0.20	0	0.28
North Atlantic	+0.46	0.15	0	0.16	+0.25	0.16	0	0.33
Europe	+0.20	0.15	0	0.55	-0.44	0.17	0	0.20
UT								
North America	-0.92	0.17	1	0.04	-1.08	0.18	1	0.01
North Atlantic	+0.38	0.18	0	0.20	+0.22	0.20	0	0.52
Europe	-0.24	0.14	0	0.73	-0.59	0.15	0	0.26
Central Asia	+0.66	0.33	0	0.53	+0.32	0.33	0	0.85
Tropical Asia	-0.58	0.39	0	0.72	-0.54	0.04	0	0.63

5
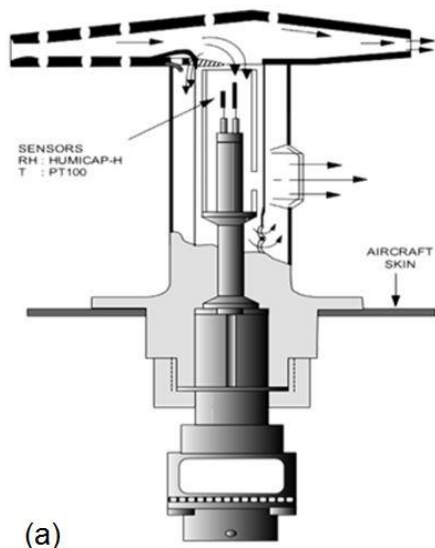
Figure 1: a) Schematic scheme of the temperature sensor attached to the humidity sensor mounted in the Rosemount housing (Helten et al., 1998). b) Packages 1 and 2 installed aboard the AIRBUS A340-300, and the inlet plate including the Rosemount housing (photograph by courtesy of Lufthansa).
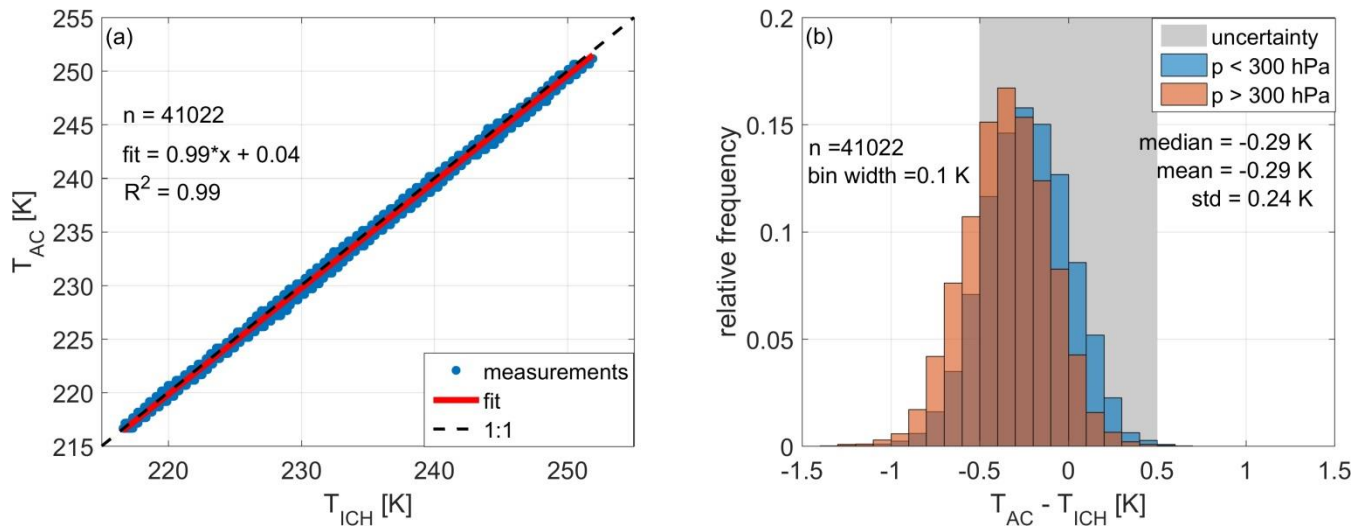
5

Figure 2: (a) Temperature correlation for all measurements below the pressure altitude 400 hPa and **(b)** Temperature bias of the IAGOS temperature instrument (T_{ICH}) compared to a regularly quality checked temperature sensor (T_{AC}) from the research aircraft for seven flights during AIRTOSS-ICE campaign in 2013. The grey area marks the total range of the total uncertainty for the IAGOS temperature observations. The temperature bias was separated for measurements between 300 and 400 hPa (orange) and 200 to 300 hPa (blue) to highlight the shift to a smaller bias at lower pressure.
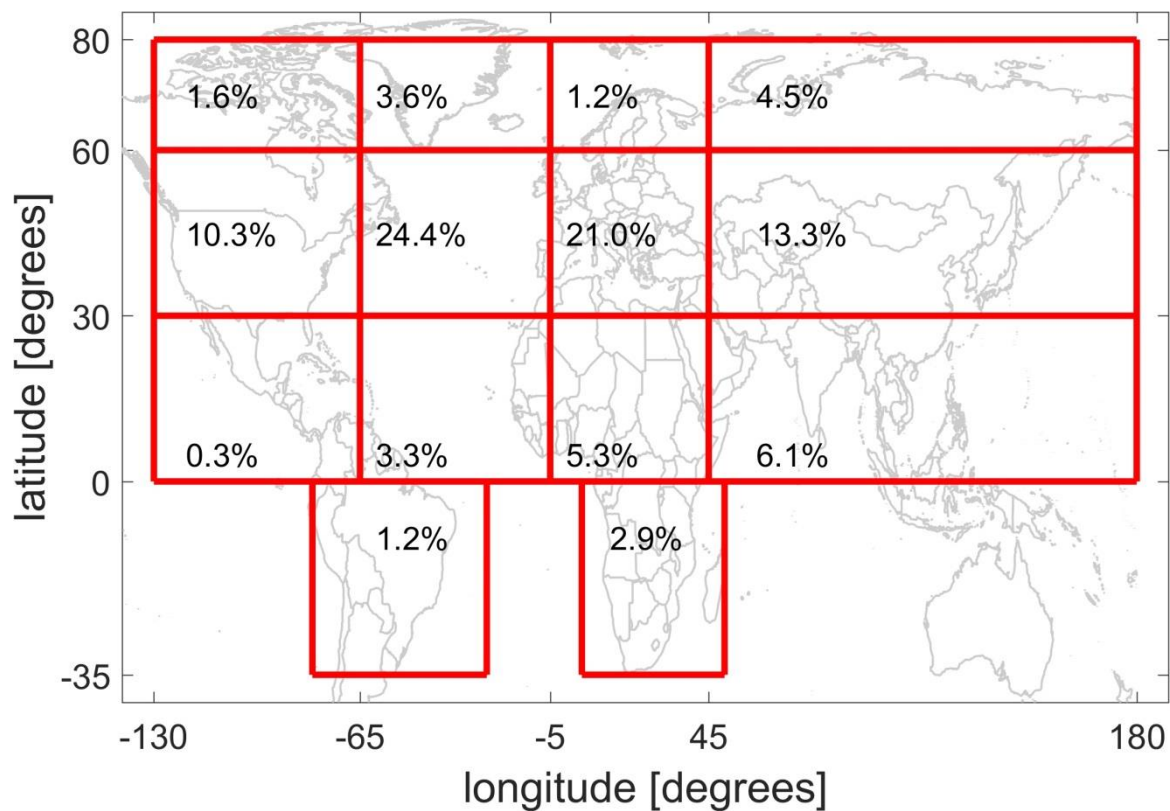
Figure 3: Relative frequency of the IAGOS temperature observations at cruise altitude (p < 350 hPa) in different regions (Tab. 1). The total number of 1-minute averages is 14.8 million. Note, 1% includes 0.2 Million 1-minute mean data points within the ULTS region
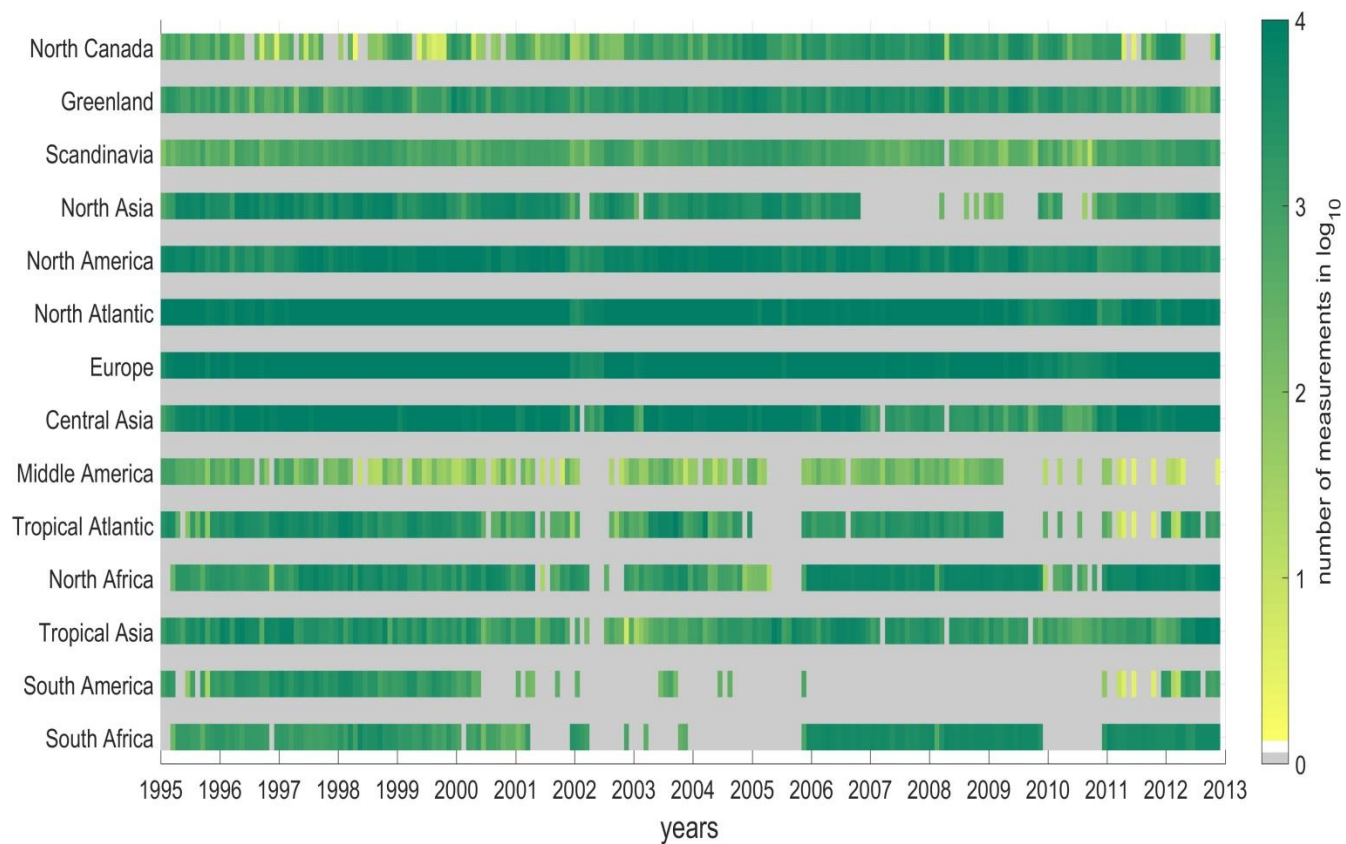
Figure 4: Data density at cruise altitude (p < 350 hPa) for all defined regions in a logarithmic scale (color) for the analyzed period.
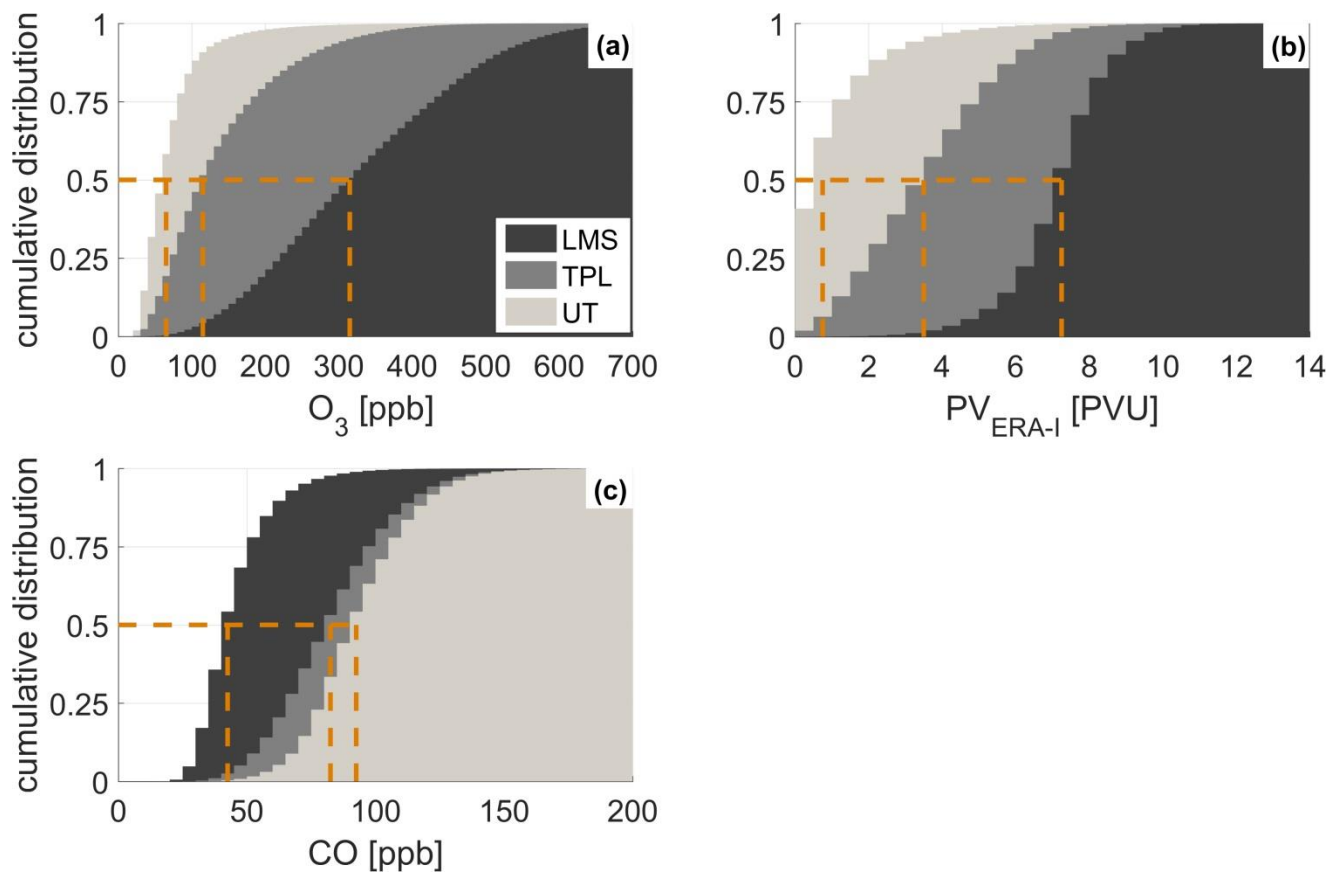
Figure 5: Cumulative distribution of a) ozone, b) PV from ERA-I, and c) CO for the different vertical layers (LMS (dark), TPL (grey), UT (light grey)) over the North Atlantic region from January 1995 to December 2012. The dashed lines mark the values of each species when the cumulative distribution reached 50%.
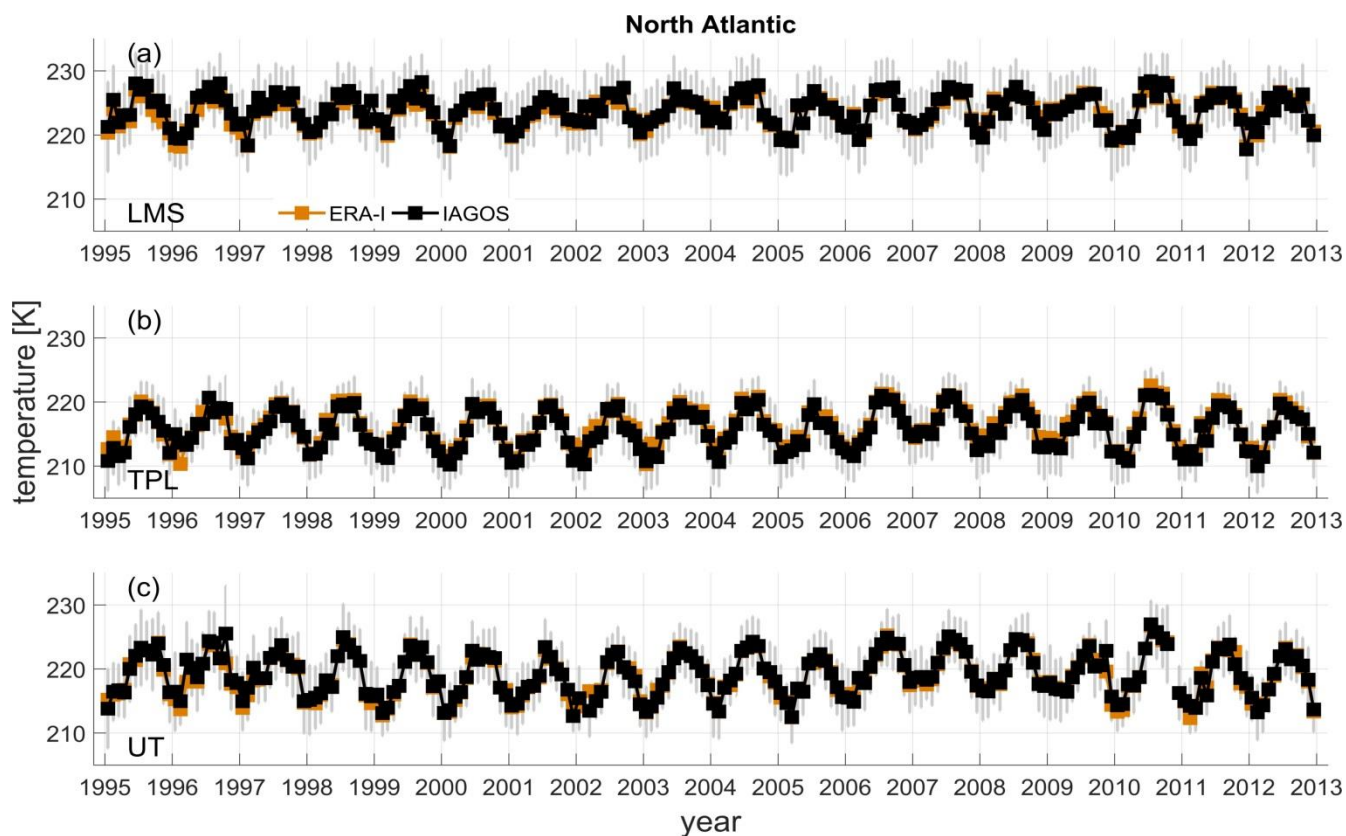
Figure 6: Temperature time series of IAGOS observations (black) and ERA-I (orange) for the a) lowermost stratosphere, b) tropopause layer, and c) upper troposphere over the North Atlantic region. The grey lines show the standard deviation (1σ) of the mean using the IAGOS observations for each month.
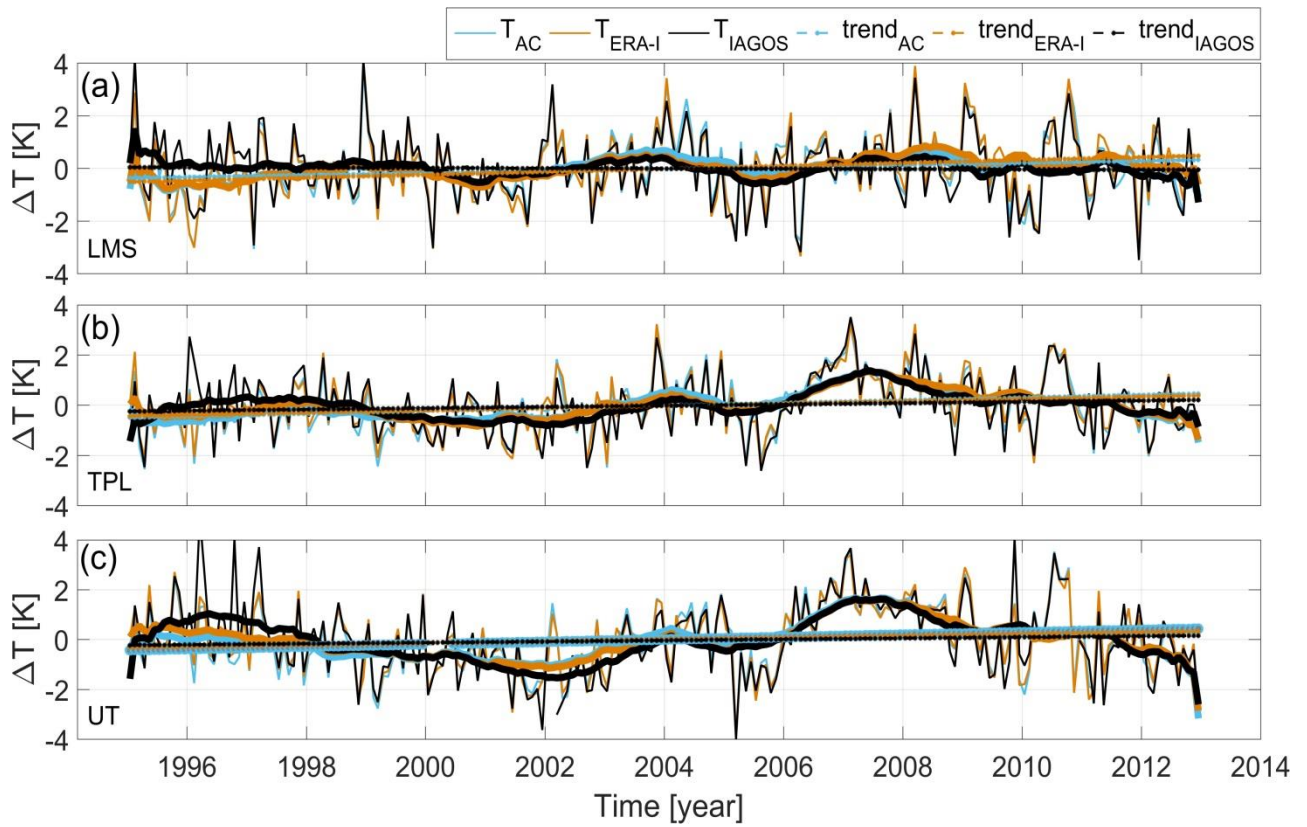
Figure 7: Anomalies in monthly averaged temperature in the a) LMS b) TPL, and c) UT over the North Atlantic region from IAGOS observations (black), ERA-I (orange), and aircraft measurements (AC, light blue). The aircraft measurements are assumed to be comparable to AMDAR data. The anomalies are smoothed with a 12 month running mean to highlight the behavior of the time series, which reflects the temperature trends by linear regression analyses (dashed lines).
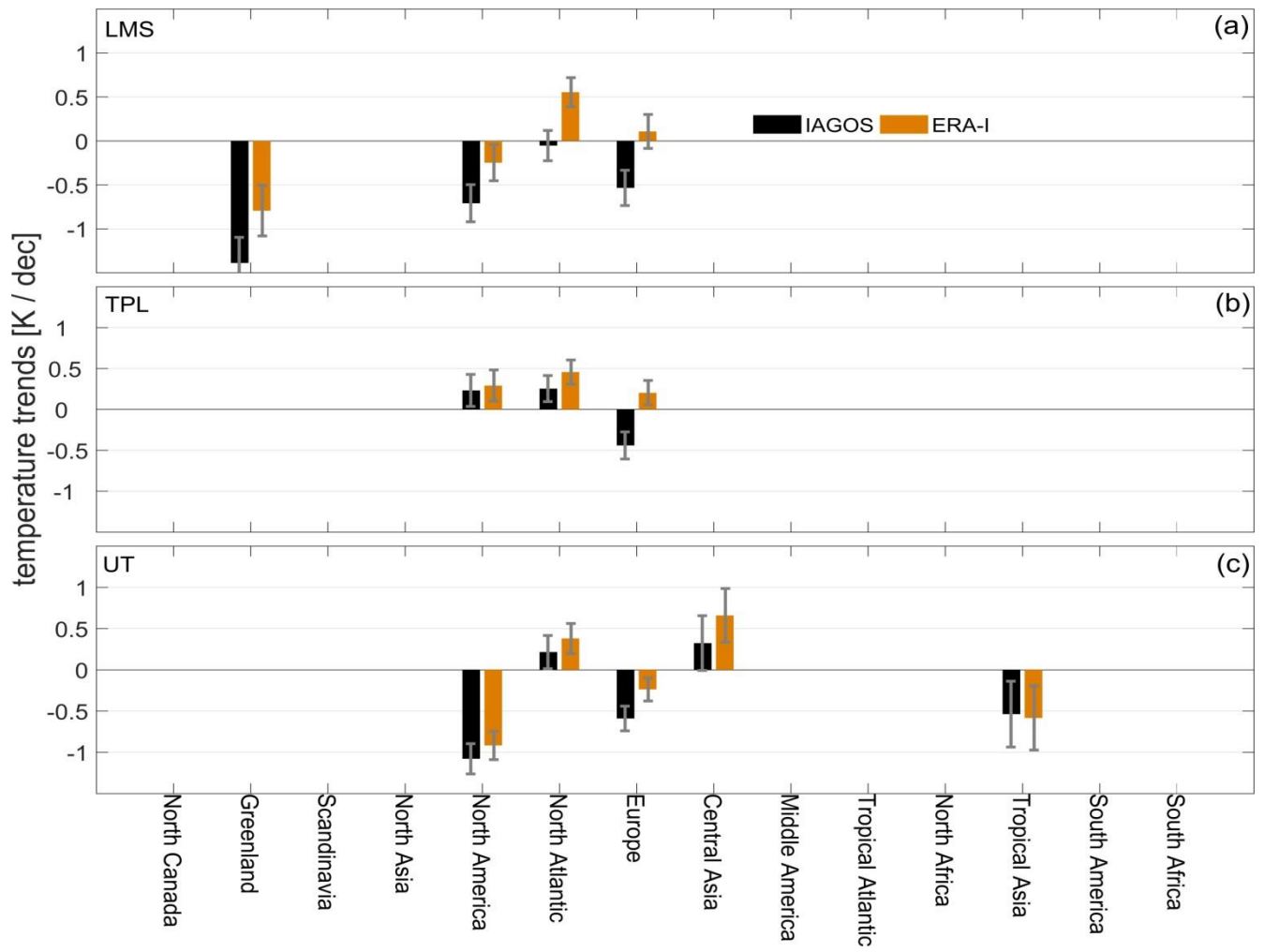
Figure 8: Temperature trends over the analyzed period for each region obtained from the IAGOS observations and ERA-I for the a) LMS, b) TPL, and c) UT, where at least 90% of all months where available (see text for details).
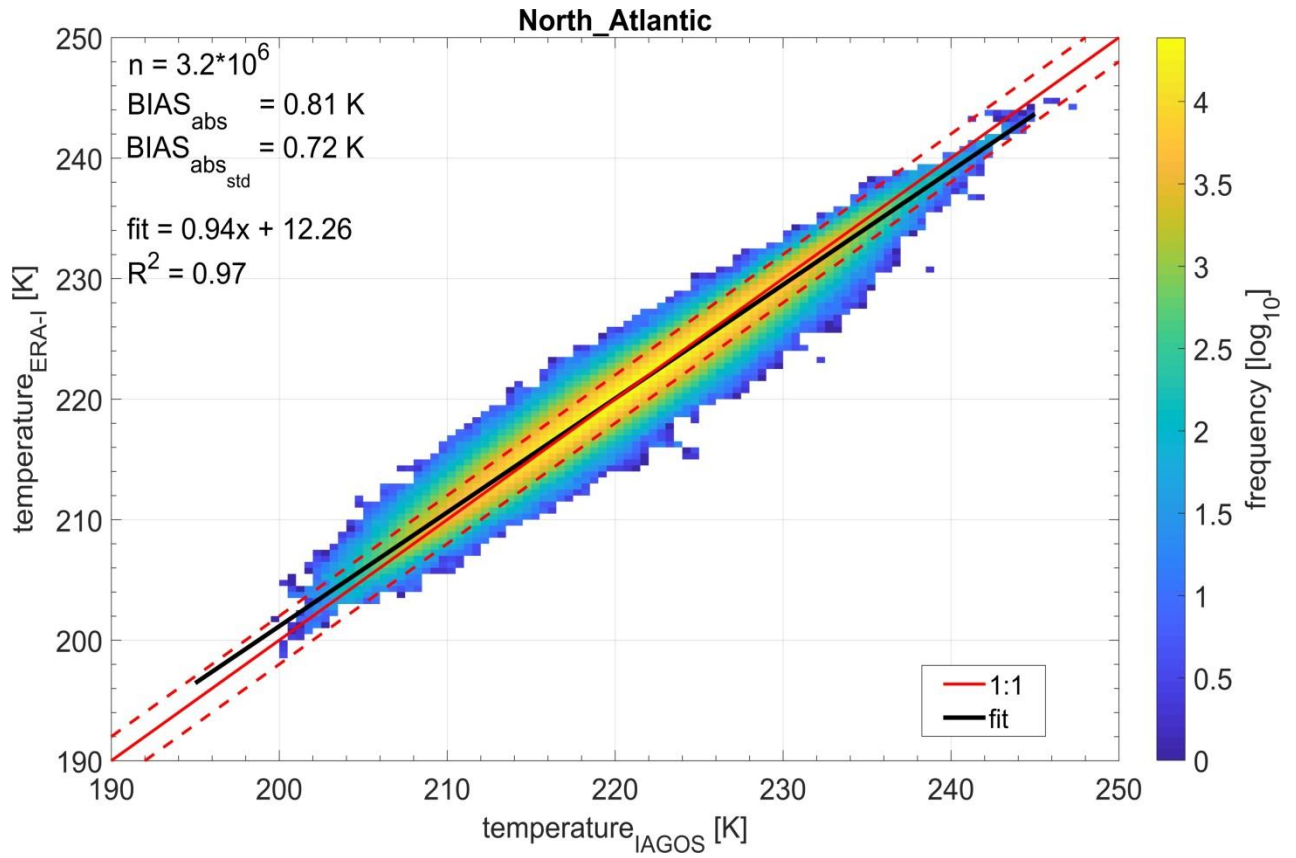
Figure 9: Temperature correlation from the observations and ERA-I over the North Atlantic region. The distribution is averaged in bin widths of 0.5 K and the color shows the number frequency with each bin in a logarithmic scale. The included numbers show the amount of data points, the absolute bias, the corresponding standard deviation and the linear robust regression fit (black line). The 1:1 line is shown in red, and its variation with 2 K as red dashes lines. The statistical values for all other regions are summarized in Tab. 3.
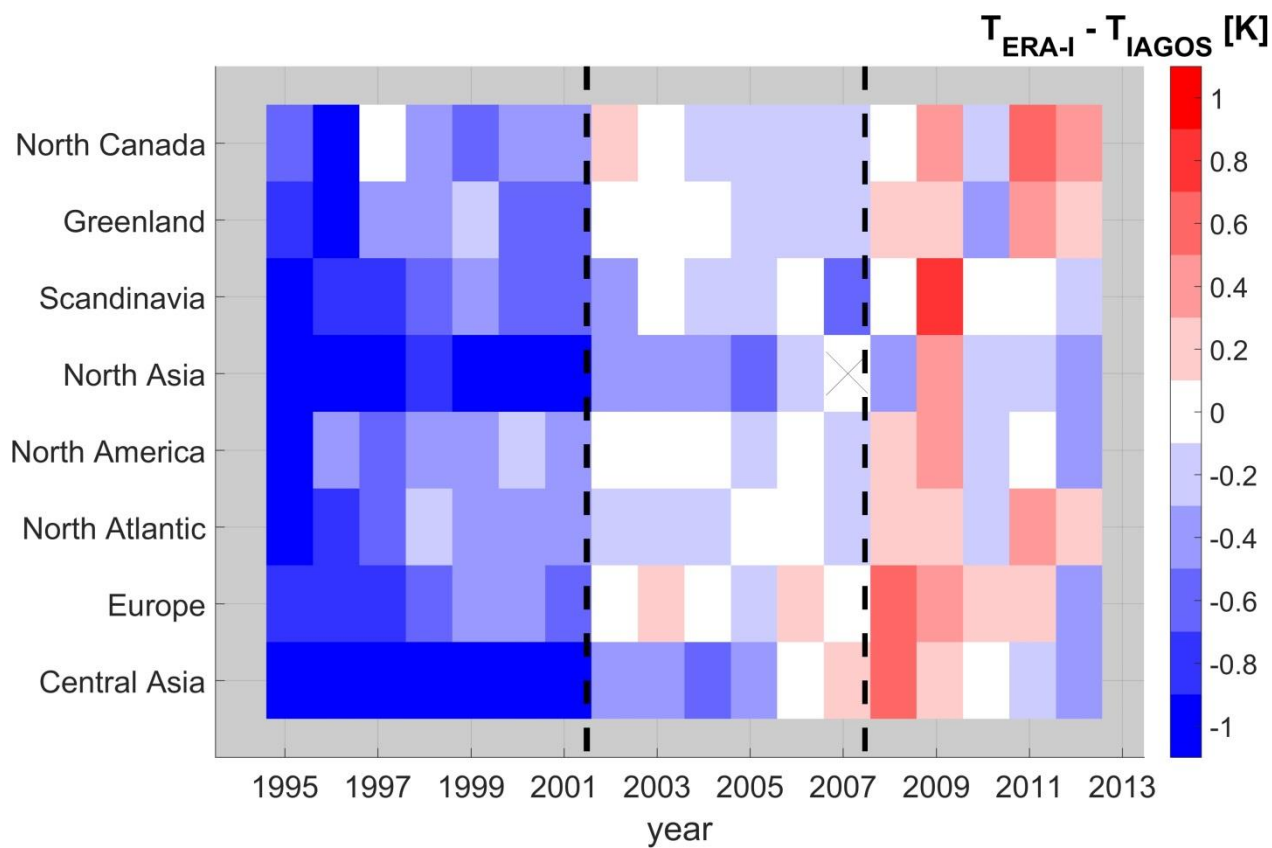
Figure 10: Annual mean of the monthly mean difference between the observations and ERA-I for different region in the Northern Hemisphere in the lowermost stratosphere from 1995 to 2014. The warm colors corresponds to colder temperatures of the observations and blueish colors to warmer IAGOS temperatures compared to ERA-I. The dashed lines show clear break-points within the time series. The cross marks a year when the annual mean could not be calculated.

Reply to

Interactive comment on "Two decades of in-situ temperature measurements in the upper troposphere and lowermost stratosphere from IAGOS long-term routine observation" by Florian Berkes et al.

Anonymous Referee #1

Received and published: 22 July 2017

This study presents a very valuable data set of 18 years of IAGOS temperature measurements in the UTLS region. Trend estimates are derived and compared with ERA-Interim reanalyses. The results are relevant and novel, and the paper is generally well written. However, a few very important aspects of the study are not clear and require clarification. I therefore recommend to accept the paper subject to major revisions.

We thank the referee for her/his comments, which we address (in bold) point by point in our reply below.

Major comments:

A) An 18-year time period is short for trend analyses. I still think that the trend calculations in this study are useful, and I very much appreciate the efforts of the author team to produce such a long high-quality data set, but the time period issue must be discussed in the paper. I would like to see two additional trend calculations: (i) for IAGOS, how do the trends change if you skip the first or last year of your 18-year time series? Such a "sensitivity test" could be interesting to assess the robustness of trend values from 18 years compared to 17 years. The additional values could be included in Fig.8.

Answer: The robustness of the trend analysis was tested with the Mann-Kendall test using different lengths of the time series. In order to further clarify this point, we have now added one sentence in section 3.2 and updated table S4 in the supplement:

"The robustness of temperature trends was tested by skipping the first or last year of the 18-year period. Within all layers and all regions each trend keeps the same sign and the trend values varied within the standard error. The only exception was the upper troposphere over North America where the temperature trend changed from slight positive trend (18 years) to neutral when the final year was removed. "

(ii) for ERA-Interim, you could compare the trends for the 18-year period with the full ERA-Interim period. Again, this would tell us something about how trends depend on the duration of the time period considered. A brief discussion of this should also be included in the concluding section.

Answer: We agree that the full time period would give additional information about the behavior of the ERA-Interim data, but the ERA-I temperature data is interpolated along the flight tracks, therefore it covers only the IAGOS period. We could create artificial flight tracks for other years, however the focus of this study is to present the IAGOS temperature measurements and not to evaluate ERA-I.

B) I appreciate the efforts in data calibration, but I find it a bit disappointing that IAGOS data is only available until December 2012. Adding some of the recent years would also help with the issue mentioned above. Is there no way how you could include a few more years?

Answer: Between 2011 and 2014 some MOZAIC aircraft were still operating. The first IAGOS aircraft started to measure in July 2011 and the second in June 2012. The IAGOS temperature and humidity data validation tool had to be re-developed and is currently under evaluation for measurements past 2013. The data coverage of the MOZAIC aircrafts for the period past 2013 is too sparse for a meaningful extension of our trend analysis. Therefore we limited our data analysis to the period until 2012.

C) Unfortunately, I don't understand the method to distinguish between LMS, TPL and UT. I understand how you determine the pressure of the thermal TP from ERA-Interim; so the TPL is a 30-hPa deep tropopause-following layer, which varies in space and time (is this correct)? Then I am lost what "max(TPL)" and "min(TPL)" mean on p. 5: max and min over what? time or space? and how to you measure max/min? does it refer to pressure? It seems to me as if TPL is tropopause-following, but LMS and UT have fixed horizontal bounds, I find this very confusing.

Answer: We are sorry for the confusion. TPL is related to the range of the pressure altitude of the tropopause +-15 hPa. Max/min is related to the upper/lower limit of this range around the tropopause. To clarify, all three layers are tropopause-following and we improved the layer definition in the manuscript:

LMS : $p < p_{TPHWMO} - 15$ hPa, which is limited by the maximum cruise altitude (p ~ 190 hPa)

TPL: $p = p_{TPHWMO} \pm 15$ hPa

UT: $p > p_{TPHWMO} + 15$ hPa, limited to 350 hPa

D) p. 6 lines 13-15: It is very important whether IAGOS data has been assimilated in ERA-Interim, or not. This is not clear from the text. The first sentence says that "aircraft and other" data are assimilated in ERA-Interim, it seems that this does not include IAGOS. The next sentence then says "Note that IAGOS ... observations are not assimilated in any other NWP model ..." which sounds as if IAGOS is assimilated only in ERA-Interim, but I assume it is not assimilated by any reanalysis system. Then the "other" would be very misleading. Please clarify.

Answer: IAGOS temperature measurements are not assimilated into ERA-Interim or any other model. We deleted the word "other" and apologize for the confusion.

Minor comments:

1) p. 1 line 21: "temperature bias between observation and model data" sounds strange to me; do you mean the bias of the observations or the bias of the "model data" (note that reanalyses are not really just model data) or do you mean that both have a bias but that the biases differ?

Answer: We agree with the reviewer that it is not very clear. We exchanged the word "bias" by "difference".

2) p. 1 line 28: delete ";" after reference to Seidel et al., 2016

Answer: Done

3) p. 2 line 7: "suffers" –> "suffer"

Answer: Changed

4) p. 2 line 15: first introduce abbreviation for LMS

Answer: Done

5) p. 2 line 29: "Petersen et al., 2015" should read "Petersen, 2016" (single author, and paper appeared one year later)

Answer: Sorry for the mistake, the reference is updated.

6) p. 2 line 30: "including" is strange here, do you mean "assimilating"?

Answer: Changed

7) p. 2 line 31: "Drue" –> "Drüe"

Answer: Changed

8) p. 3 line 12: not clear what is meant by "both types of profile measurements"

Answer: Should read: *"… with temperature profiles from in-situ measurements and from the model…"*

9) p. 4 line 22: most readers are not familiar with the AIRTOSS-ICE campaign. What "research aircraft" has been used, and what type of temperature sensor?

Answer: We included the information in the text

"During this campaign, the IAGOS temperature instrument was installed on the research aircraft (Learjet 35A), and provides the opportunity to compare both temperature measurements. The aircraft temperature measurement was made with Pt-100 thermistor mounted in the same type of Rosemount as for the IAGOS temperature measurements. The temperature sensor of the research aircraft has been calibrated regularly with an uncertainty of about 0.5 to 1.0 K."

10) p. 4 line 31: I understand "air temperature" but what is "total air temperature"?

Answer: We included into the text:

"The total air temperature is defined as the ambient air temperature plus the temperature increase due to adiabatic compression in the Rosemount housing."

11) p. 5 line 3: "maintained" sounds strange, maybe "made"?

Answer: We changed *"maintained"* to *"performed"*.

12) p. 5 line 4: I find this a strange remark, it sounds as if the DWD model was used as a reference to test the accuracy of the observations … maybe delete this sentence.

Answer: We agree with the reviewer that the sentence leads to misunderstanding and deleted it.

13) p. 5 line 21 and in other places: use "p" as symbol for pressure, not "P"

Answer: Changed

14) p. 5 line 30: it is fine that you use here the thermal tropopause, but the long list of chemical,…tropopauses does not help the reader and is not relevant here. I suggest to delete the last 3.5 lines on this page.

Answer: We follow the suggestion of the reviewer and delete the different tropopause definitions.

15) p. 6 line 2: of course the Coriolis parameter is defined near the equator, but it is zero! Therefore PV goes to zero and the 2-pvu surface is not defined.

Answer: Thanks for the correction. We changed the sentence to:

The thermal tropopause is valid within all latitude bands, whereas the dynamical tropopause cannot be used in tropical regions, because the Coriolis parameter is zero, therefore PV goes to zero and the 2-pvu surface is not defined (Boothe and Homeyer, 2017).

16) p. 6 line 19: Reichler et al. 2003 is not in the list of references

Answer: Sorry for the mistake. It is now included.

17) p. 6 line 21: this statement is very strange, the time resolution of ERA-Interim is 6 hours, so it cannot be reduced to 1 minute. Please explain how you interpolated 1-min values from ERA-Interim, by linear interpolation in time?

Answer: The ERA-I data is linearly interpolated (time and space) along the flight path in 4 s temporal resolution equivalent to the in-situ measurements. Afterwards the in-situ data and the ERA-I data were averaged to 1 min to reduce the size of data as written on p. 4 line: 19.

18) p. 8 line 3: "it is assumed …": do you have a reference for this?

Answer: We included the reference *"Kandhu et al. 2016"* at the end of the sentence:

Khandu, Awange, J. L. and Forootan, E.: Interannual variability of temperature in the UTLS region over Ganges–Brahmaputra–Meghna river basin based on COSMIC GNSS RO data, Atmos. Meas. Tech., 9(4), 1685–1699, doi:10.5194/amt-9-1685-2016, 2016.

19) p. 8 line 6: "in this region" appears twice in the same sentence

Answer: We rephrased the sentence.

"Another reason could be related to a higher variability of the temperature due to large scale influence or simply that the data coverage is still too poor in this region, which might lead to a higher variability of the local temperatures which then mask the temperature trend."

20) p. 8 line 12: what do you mean by "local tropopause", maybe delete "local"?

Answer: We agree with the reviewer and deleted local.

21) p. 8 line 30: "where" –> "when"

Answer: Changed

22) p. 9 line 21: "then" –> "than"

Answer: Changed

23) p. 9 line 26: "annual averages of the monthly deviation": sounds complicated to me and it should be the same as just "annual deviation"?

Answer: Changed as suggested

24) p. 10 line 16: references should be in chronological order

Answer: Corrected

25) p. 10 line 19: I don't understand this sentence: what aircraft measurements are you using here? IAGOS or AMDAR, and what is assumed to be similar?

Answer: We demonstrated that using only the aircraft temperature measurements ("AMDAR like") the aircraft temperature trend follow mostly the ERA-Interim temperature trend, which is an indication that here is a positive bias in the aircraft data. However, the aircraft temperature data were not validated after the AMDAR recommendations, therefore we wrote: "…, which we assume to be comparable to AMDAR measurements". We included the "(T_{AC})" to be more specific.

26) p. 11 line 20: what is meant by "use IAGOS observations as anchor point"? Do you suggest to calibrate AMDAR data with IAGOS? And why do you not suggest that IAGOS observations should be assimilated in, e.g., ERA-5?

Answer: The amount of IAGOS temperature measurements are much less compared to the amount of AMDAR temperature measurements and the benefits would be minor in respect to additional measurements. However, both temperature measurements (AMDAR and IAGOS) could be assimilated into ERA5, than IAGOS data could be uses as anchor point for the bias correction of the AMDAR data.

27) p. 15 line 13: I did not find a reference to Kuo et al. in the text

Answer: Sorry for the mistake. The reference is an artefact from an earlier version of the manuscript.

28) p. 16 line 9: volume and page numbers are missing

Answer: Fixed

Table S4: Temperature trends of ERA-I and from the IAGOS observations within the LMS, TPL and UT as shown in table3 and compared to temperature trends derived from 17 years skipping the first (light gray) or last year (dark gray) in the analyses.

Region	ERA-I				IAGOS			
	ΔT_{18yr} K/dec	SE K/dec	$\Delta T_{17yr, first}$ K/dec	$\Delta T_{17yr, last}$ K/dec	ΔT_{18yr} K/dec	SE K/dec	$\Delta T_{17yr, first}$ K/dec	$\Delta T_{17yr, last}$ K/dec
LMS								
Greenland	-0.79	0.29	-1.03	-0.83	-1.39	0.29	-1.48	-1.45
North America	-0.25	0.21	-0.38	-0.21	-0.71	0.21	-0.73	-0.76
North Atlantic	+0.56	0.17	+0.53	+0.59	-0.05	0.17	+0.02	-0.01
Europe	+0.11	0.19	+0.12	+0.14	-0.53	0.20	-0.49	-0.62
TPL								
North America	+0.29	0.19	+0.11	+0.45	+0.23	0.20	-0.02	+0.42
North Atlantic	+0.46	0.15	+0.42	+0.62	+0.25	0.16	+0.20	+0.38
Europe	+0.20	0.15	+0.17	+0.19	-0.44	0.17	-0.46	-0.57
UT								
North America	-0.92	0.17	-0.99	-0.82	-1.08	0.18	-1.16	-1.00
North Atlantic	+0.38	0.18	+0.52	+0.58	+0.22	0.20	+0.33	+0.43
Europe	-0.24	0.14	-0.21	-0.29	-0.59	0.15	-0.55	-0.71
Central Asia	+0.66	0.33	+0.55	+0.91	+0.32	0.33	+0.27	+0.54
Tropical Asia	-0.58	0.39	-0.43	-0.24	-0.54	0.04	-0.30	-0.21

Reply to

Interactive comment on "Two decades of in-situ temperature measurements in the upper troposphere and lowermost stratosphere from IAGOS long-term routine observation" by Florian Berkes et al.

Anonymous Referee #2

Received and published: 28 July 2017

This paper analyzes passenger-aircraft IAGOS temperature observations in the UTLS for the period 1995-2012, focusing on long-term trends. Comparisons with the ERA-Interim reanalysis data are also made, and change points in ERA-Interim are identified and discussed. Very careful data analysis has been made, and the discussion on ERA-Interim is very interesting. I think the manuscript can be published in ACP after considering the following comments.

We thank the referee for her/his comments, which we address (in bold) point by point in our reply below.

1. As the authors explain in Introduction, the aircraft AMDAR data are known to have warm biases. I would like the authors to add some more explanations why the IAGOS data are considered to be without such biases. Specifically, please consider

(1) to add a similar figure to Figure 1 for the automated temperature instrument and

<u>Answer:</u> **The aircraft temperature sensor position and Rosemount housing are located at three different spots at the nose region of the aircraft (may differ from airline to airline), but the system is equivalent to the IAGOS system and this is shown.**

(2) to add some technical descriptions for it in the last paragraph of Section 2.2. Is the temperature sensor material same? Is the instrument housing same? Is the correction method for adiabatic compression same? Are there other data processing procedures for both?

<u>Answer:</u> **Moninger et al. (2003) summarized that the aircraft temperature sensor can be affected by moisture, and dirt (insects and other materials), which can lead to a coating of the probe and to a drift of the sensor signal. The sensor type is also a platinum resistance sensor, as used in IAGOS. Temperature measurement uncertainties increase due to the adiabatic temperature correction, which includes the uncertainties of the static pressure and Mach number (calculated from the aircraft speed). The aircraft temperature correction of the adiabatic compression is calculated by the avionics of the aircraft itself. The entire procedures and uncertainty determination are given in Stickney et al., 1994 and Helten et al., 1998. The aircraft temperature is provided from avionics of the aircraft with a precision of +/- 0.5 K and maximum uncertainty of +/- 2 K based on AMDAR quality criteria (WMO, 2003). Please note, we wrote in the introduction of the manuscript: "** However, Ballish and Kumar (2008) and Drüe et al. (2008) identified that the AMDAR aircraft temperature is strongly affected by a warm bias, which can fluctuate by altitude, aircraft type and phase of flight, while the reason for this bias is not fully understood (Ingleby et al., 2016)."

IAGOS changes regularly (currently 2-3 months interval) the temperature/humidity sensors to avoid drifts of the signal (e.g. due to dirt) and the quality of each sensor is checked before and after each deployment in the laboratory. Also for each 4-s data point we provide a quality flag and the uncertainty is calculated, which is not the case for the aircraft temperature data. As far as we know, AMDAR temperature measurements aren't considered if they differ more than 2 K from the NWPs. So we think that we might be able to avoid a warm temperature bias, because we have a more frequent control and exchange of the temperature sensor.

We included additional information in the last paragraph of Sec 2.2 in the revised version.

2. Please also describe the instruments for positions, pressure, and aircraft and wind speed/direction? The pressure measurement uncertainty may be important for analyzing temperature trends because there are temperature gradients in the UTLS so that pressure errors would result in slightly different temperature values. Aircraft/wind speed/direction might be important for the correction of adiabatic compression (or not?).

<u>Answer:</u> **The instruments might vary for each aircraft type, a general overview about the uncertainty of the aircraft data within the IAGOS project was given by Petzold et al., 2015. It is true that the aircraft speed is crucial for the correction of the adiabatic compression, which is applied for each individual aircraft based on the in-situ measurements. The correction factor is calculated using the following equation:**

$T_{total} / T_{ambient} = (1 + (\gamma-1)/2) M^2)$

With T_{total} as the total air temperature measured from the sensor. γ is the ratio of the specific heats of air at constant pressure and volume, respectively. M is the Mach number which is calculated using the aircraft speed divided by speed of sound in air. The total uncertainty of the temperature measurements is calculated with 0.5 K and includes the temperature sensor uncertainty of 0.25 K plus the uncertainty for the correction of the adiabatic compression (Helten et al., 1998). At cruise altitude, the Mach number is rather constant at M= 0.81+-0.01. More details are given in the standard operation procedure (SOP) of the IAGOS capacitive Hygrometer (ICH), available at www.iagos.org.

The aircraft pressure measurements are used to derive the aircraft position relative to the thermal tropopause calculated from ERA-Interim. The uncertainty from the pressure sensor (0.35 hPa) which is small compared to our range of the tropopause layer (+/- 15 hPa). Note, we wrote in the manuscript, that we evaluate the vertical position of the aircraft using in-situ measurements of CO, O3 and PV.

We added additional information at different parts in the manuscript.

 3. I have some questions related to the data analysis methods. (Please note that I am not asking the authors to re-do the data analysis.)

3-1. Are there any sampling biases/trends in vertical within each layer? Because there are temperature gradients in the UTLS, if the flight level would have changed systematically over the period, we would get apparent temperature trends.

3-2. Similarly, are there any sampling trends in horizontal? My understanding is that the pilots are trying to avoid turbulences i.e., the regions with strong (horizontal) shears. If the westerly jets have some trends in their locations (cf. Davis and Rosenlof, 2012; Williams, 2017), the aircraft paths may have some systematic changes that could give apparent temperature trends.

Answer: We looked at the beginning in detail at the horizontal and vertical data coverage for each specific region, and we did not find a shift of the data coverage within each region over the study period. Over most regions the aircraft fly along distinct routes. Over the North Atlantic and Tropical Atlantic regions, the spread of the fight tracks is larger, which corresponds to flight routes with different destinations and avoiding severe weather conditions. However, we agree with the reviewer that it would be worth to look more into details of the aircraft measured wind fields to study changes of the wind pattern, which, however, is not the focus of this study.

3-3. For comparison with ERA-Interim, another way could be to pick up 6 hourly data (i.e., at the times when the reanalysis data are provided) from the IAGOS data. Linear interpolations to 1 min from 6 hourly data may produce unnecessary spreads in the difference of the two data.

Answer: The ERA-Interim data is interpolated along the flight path for every 4s and then averaged to 1 min to reduce the amount of data. If we compare point by point, the spread is large between the temperature of ERA-I and in-situ as we wrote in the manuscript and show in Fig 9. However over 93% of this spread is smaller than 2 K for the North Atlantic, which reflects the generally good agreement between these two data sets. Of course the IAGOS data could be interpolated on a model grid with the same time resolution, but from our point of view, this would lead to a new study with the focus on the evaluation the ERA-Interim model. For such study, it would be favorable to include additional data sets (e.g MERRA2, JRA-55 and others), which in fact is subject of ongoing work.

4. Page 7, lines 23-24 and Figure 7. The warm phase during 2006-2008 might be related to a volcanic eruption (the Soufrière Hills at 16N on 20 May 2006). See Vernier et al. (2011).

Answer: The reviewer is right that the mentioned warm phase could be related to the volcanic eruption. However, this hypothesis is speculative and we decided to avoid it and leave it for a future study.

References:

Sean M. Davis, S. M., K. H. Rosenlof (2012), A Multidiagnostic Intercomparison of Tropical-Width Time Series Using Reanalyses and Satellite Observations, J. Clim., 25, 1061-1078, https://doi.org/10.1175/JCLI-D-11-00127.1.

Vernier, J.-P., et al. (2011), Major influence of tropical volcanic eruptions on the stratospheric aerosol layer during the last decade, Geophys. Res. Lett., 38, L12807, doi:10.1029/2011GL047563.

Williams, P.D., Increased light, moderate, and severe clear-air turbulence in response to climate change, Adv. Atmos. Sci. (2017) 34: 576, https://doi.org/10.1007/s00376-017-6268-2.