



RESEARCH ARTICLE

Sensitivity of trends to estimation methods and quantification of subsampling effects in global radiosounding temperature and humidity time series

Souleymane SY  | Fabio Madonna  | Marco Rosoldi |
Emanuele Tramutola | Simone Gagliardi | Monica Proto | Gelsomina Pappalardo

Consiglio Nazionale delle Ricerche,
Istituto di Metodologie per l'Analisi
Ambientale (CNR-IMAA), Potenza, Italy

Correspondence

Souleymane SY and Fabio Madonna,
Consiglio Nazionale delle Ricerche (CNR),
Istituto di Metodologie per l'Analisi
Ambientale (CNR-IMAA), Zona
Industriale, 85050 Tito Scalo (PZ),
Potenza, Italy.

Email: souleymane.sy@imaa.cnr.it (S. S.)
and fabio.madonna@imaa.cnr.it (F. M.)

Funding information

European Union's Copernicus Climate
Change Service (C3S) implemented by
European Centre for Medium-Range
Weather Forecasts (ECMWF), Grant/
Award Number: 2017/
C3S_311a_Lot3_CNR/SC2

Abstract

Climate trends estimated using historical radiosounding time series may be significantly affected by the choice of the regression method to use, as well as by a subsampling of the dataset often adopted in specific applications. These are contributions to the uncertainty of trend estimations, which have been quantified in literature, although on specific pairs of regression methods, and in the not very recent past characterized by smaller trends in temperature than those observed over the last two decades. This paper investigates the sensitivity of trend estimations to four linear regression methods (parametric and nonparametric) and to the artificial subsampling of the same dataset using historical radiosounding time series from 1978 onwards, available in the version 2 of the Integrated Global Radiosonde Archive (IGRA). Results show that long-term decadal trends may have not negligible uncertainties related to the choice of the regression method, the percentage of data available, the amount of missing data and the number of stations selected in the dataset. The choice of the regression methods increases uncertainties in the decadal trends ranging from -0.10 to -0.01 $\text{K}\cdot\text{da}^{-1}$ for temperature in the lower stratosphere at 100 hPa and from 0.2 to 0.8% da^{-1} for relative humidity (RH) in the middle troposphere at 300 hPa. Differences can also increase up to 0.4 $\text{K}\cdot\text{da}^{-1}$ at 300 hPa when the amount of missing data exceeds 50% of the original dataset for temperature, while for RH, significant differences are observed in the lower troposphere at 925 hPa for almost all datasets. Finally, subsampling effects on trend estimation are quantified by artificially reducing the size of the IGRA dataset: Results show that subsampling effects on trend estimations when at least 60 stations, up to 76% of data available, are considered for temperature and at least 40 stations for RH.

KEYWORDS

climate trends, radiosonde, regression methods, sampling effect, uncertainty

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors *International Journal of Climatology* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

1 | INTRODUCTION

Historical radiosounding observations have proven to be extremely successful for climate monitoring and for trend detection by providing high vertical resolution profiles of Essential Climate Variables (ECVs). Radiosounding profiles of temperature, humidity and wind have been used both to characterize the structure of weather systems and to detect changes in climate evolution. Radio soundings are not the only source of upper air data but offer a few potential advantages compared to satellite observing platforms. The satellite microwave sounding units (MSUs) have provided global data since 1979 with spatial coverage far exceeding that of radiosounding networks, especially in the Tropics, Southern Hemisphere (SH) and oceanic regions but with a coarse vertical resolution, especially in the upper troposphere (UT) and lower stratosphere (LT) regions (Christy and Goodridge, 1995; Gaffen, 1998; Christy *et al.*, 2003; Ladstädter *et al.*, 2011). In addition, MSU datasets are relatively short in sampling time and potentially influenced by several issues such as changes in cloud amount and tropospheric humidity (Simmons *et al.*, 2014).

At the global scale, more than 1,000 radiosoundings are approximately launched at day and night time each day to characterize the state of the atmosphere (Ladstädter *et al.*, 2011), although with large sparse coverage in the SH and occasional launches performed on ships to fill the gap over the oceans. The use of radiosounding datasets for climate change studies implies the handling of inhomogeneities caused by frequent changes in instruments, sensor types, observation practices or change in data processing systems (Thorne *et al.*, 2005; Karl *et al.*, 2006; Miloshevich *et al.*, 2006; Simmons *et al.*, 2014; Madonna *et al.*, 2020a, 2020b). Radiosondes from different manufacturers are affected by systematic errors, which differ from one type to another (e.g., Dirksen *et al.*, 2014; Ingleby, 2017).

Even though random errors in the trend estimations can be reduced by increasing the number of measurements (Weatherhead *et al.*, 1998; Whiteman *et al.*, 2011), the systematic errors persist regardless of the number of observations introducing biases in climate applications (Ingleby, 2017). In addition, changes in the utilized sensor type at different locations over time were almost not documented before 2000 and not always adequately after, leading to artificial trends or jumps in the radiosounding data records (Dai *et al.*, 2010). Finally, radiosounding observation data routinely deal with data continuity problems. In this respect, a recent study of Ferreira *et al.*, (2019) provides a detailed analysis of the completeness of radiosounding humidity observations available in IGRA version 2 (Durre *et al.*, 2018), with several statistics which

reveal that the length and completeness of humidity data vary widely among stations, and the vertical resolution, vertical extent and completeness of soundings have improved considerably over time since 1947 to the present.

Regular temporal sampling and continuity in historical radiosounding data profiles are an important aspect of trend estimation and interpretation (Schlegel and Smit, 2016). Gaffen *et al.* (2000) quantified the impact of geographical distribution and data quality on upper-air temperature trends that were estimated using National Oceanic and Atmospheric Administration (NOAA) radiosounding data for the period 1959–1995. Trend estimations were based on a comparison between the robust and nonparametric median of the pairwise slopes method and the classic parametric least-squares regression method: Results show similar trend estimations, with differences generally less than $\pm 0.03 \text{ K}\cdot\text{decade}^{-1}$ for the period 1959–1995, although differences in stratospheric trends may rarely exceed $0.10 \text{ K}\cdot\text{decade}^{-1}$ for the period 1970–1995.

Referring to different datasets obtained from MSU, radiosonde measurements and reanalysis from European Centre for Medium-Range Weather Forecasts (ECMWF) and National Centers for Environmental Prediction (NCEP), Santer *et al.* (2000) compared two linear trend estimation methods, one based on minimization of absolute deviations (LA) and another based on minimization of squared deviations (LS). Results indicate that trend estimates with LA are less sensitive to the noise than trend estimates of the LS method. They also found that differences between the two linear methods are generally less than $0.05 \text{ K}\cdot\text{decade}^{-1}$ over 1959–1996 and exceed $0.10 \text{ K}\cdot\text{decade}^{-1}$ for lower troposphere and $0.15 \text{ K}\cdot\text{decade}^{-1}$ for the lower stratosphere over 1973–1993.

The sensitivity of estimated trends to regression methods is an uncertainty contribution, which adds to the sampling uncertainty in time, due to gaps (e.g., missing data) in the data records, and in space, due to the need to select the most reliable subset of stations (Rosen *et al.*, 2003; Agudelo and Curry, 2004; Free and Seidel, 2005; McCarthy, 2008; Ladstädter *et al.*, 2011). It is worth pointing out that uncertainties due to temporal and spatial gaps in historical radiosounding data records have been often neglected in literature or reduced by using temporal averaging, on monthly or yearly scales, or data gridding using the interpolation or extrapolation method, which can reduce sampling bias but not remove it completely (McCarthy, 2008).

From NCEP reanalysis and MSU satellite data and from seven radiosonde networks available for the period 1960–1997, Free and Seidel (2005) found that sampling errors affecting trend estimation depend upon the

atmospheric pressure level and variability, the network size and the climate region. Results found that sampling errors range from less than $0.001 \text{ K-decade}^{-1}$ to more than $0.10 \text{ K-decade}^{-1}$ and differ noticeably from one dataset to another. Comparing datasets from GPS radio occultation and radiosounding observations, Ladstädter *et al.* (2011) also found that sampling errors from radiosounding datasets are generally small (less than 0.3 K) between 50°S and 50°N but are larger at higher latitudes due to greater variability of the atmosphere and to the small number of stations in the SH.

Most of above-mentioned papers used radiosounding data until 2000 and, therefore, are characterized by trend rates slower than those observed over the last two decades. Moreover, beyond the paper cited above, many of the previous studies have estimated trends in temperature or humidity from historical radiosounding by using the parametric linear regression method, which is not a robust method and may be adversely affected by outliers, by non-Gaussian behaviour of the underlying data distribution and non-stationarity (Lanzante, 1996; Zaman *et al.*, 2001; Mudelsee, 2019).

The outlined scenario shows that further studies and reviews are needed to properly quantify trend estimation uncertainties. Using historical radiosounding time series available from the IGRA (version 2) for the period 1978–2018, this paper provides a quantitative analysis of the uncertainties in the estimation of decadal trends of temperature and humidity due to the use of linear regression methods. It also provides a quantitative estimation of the uncertainty introduced by the spatial and temporal subsampling effects on decadal trends using a novel approach, which may be considered useful for the design of a measurements network, as well as for climate applications.

The paper is structured as follows: Section 2 discusses the radiosounding datasets and the related available metadata provided by IGRA. Section 3 outlines data analysis, subsampling strategies and the linear regression methods applied to estimate trends. Section 4 discusses the uncertainty contributions that may derive from the use of different linear estimation methods by comparing parametric simple linear regression method and nonparametric regression methods. Section 5 investigates the sensitivity of trends due to missing data. Section 6 describes the effect of spatial subsampling on trend estimations. Finally, a discussion and recommendations for future works are provided in Section 7.

2 | RADIOSONDE DATASETS

The radiosounding time series used for the analysis presented in this paper is a part of IGRA (version 2) (Durre

et al., 2018). To date, IGRA is the largest and most comprehensive dataset of quality-assured radiosonde observations freely available. It consists of a collection of historical and near real-time radiosounding and pilot-balloon observations taken from 2,700 globally distributed stations (including ships, buoys and remote islands over the oceans) with varying periods of record, aiming to enhance spatial and temporal coverage, particularly prior to the 1970s (see also <https://www.ncdc.noaa.gov/data-access/weather-balloon/integrated-global-radiosonde-archive>). Approximately 1,000 of the IGRA stations are currently reporting data. Compared to the former release (Durre *et al.*, 2006), IGRA version 2 increased its data volume by 30%, extending the data back in time to as early as 1905 and improving the spatial coverage, for example, in South America and Africa. Among the list of stations available from the IGRA data repository, a substantial subset of globally distributed radiosounding stations (656 stations) was used in this paper (Figure 1), namely, those with metadata available since 2000 to present.

Ferreira *et al.* (2019) found that the total number, spatial distribution and temporal completeness of IGRA stations vary considerably over the years due to several factors, and the number of vertical levels and the vertical extent in radiosonde reports depend on the standard and mandatory pressure levels provided according to the World Meteorological Organization (WMO) recommendations (WMO, 1996) to the Met services, as well as on the number of reported significant levels. Finally, due to various reasons, missing data are frequently found. In this work, monthly radiosounding time series from 1978 onwards on 16 mandatory pressure levels (1,000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20 and 10 hPa) are selected for temperature, while for RH, only data at pressures higher than 300 hPa were considered as RH sensors below 300 hPa are prone to systematic uncertainties.

The data availability of measurements in different zonal regions at each pressure level from 1,000 to 10 hPa was investigated to quantify the amount of missing data and to identify the periods where they are more frequent. Missing data are naturally present in the selected dataset due to a lack of continuity in the operations that contingently occurs at each station. This implies that, to provide a reliable assessment of the uncertainties affecting the estimation of trends using radiosounding data, the dataset selected must be investigated without restrictions on where the missing data are located, although it is known that these may affect trend estimates.

Figure 1 shows the spatial distribution of radiosonde stations with an amount of missing monthly data equivalent to 0, 5, 10, 20 and 30 years, reported as the percentage of available data with respect to the expected full

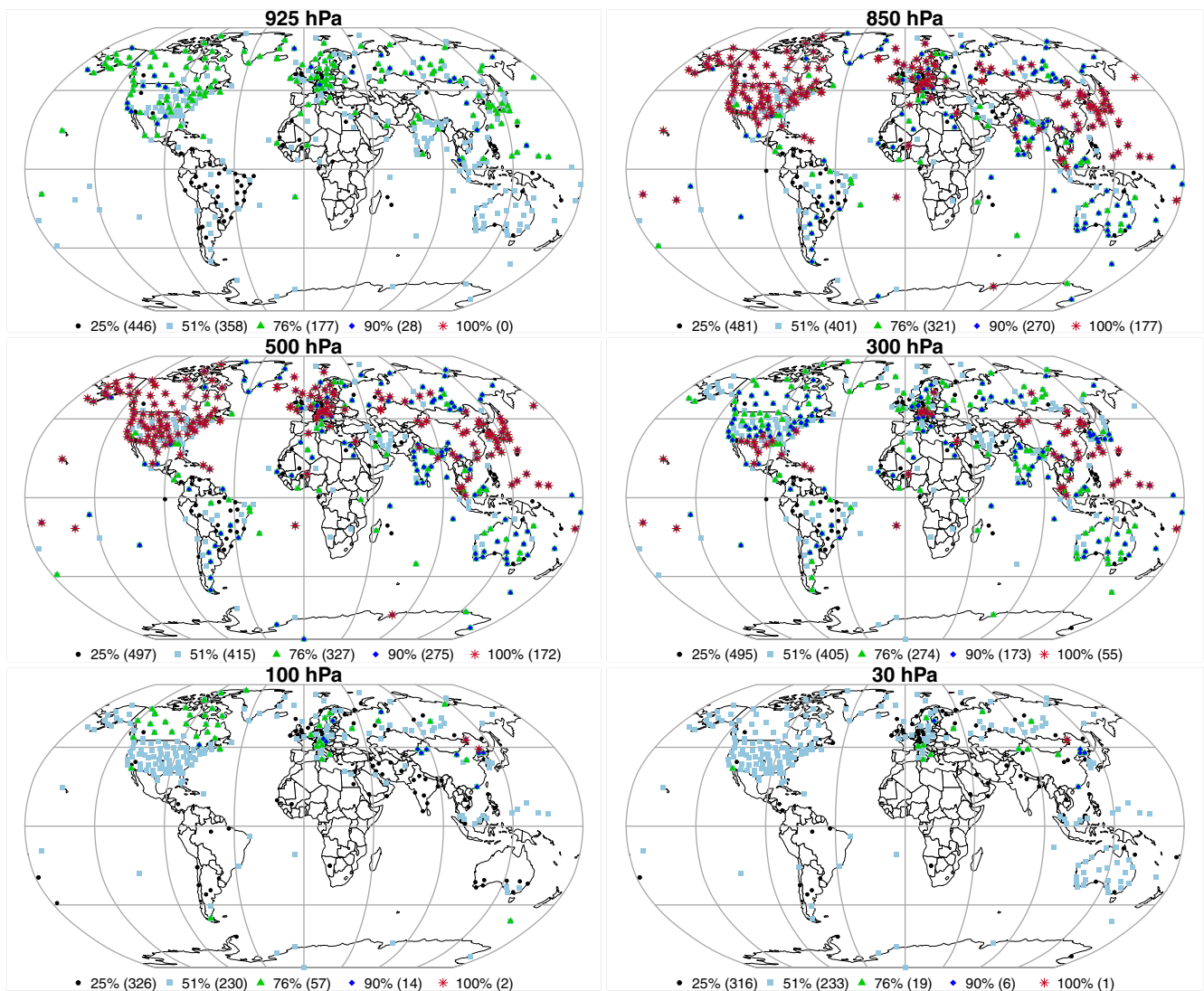


FIGURE 1 Spatial distribution of radiosounding stations with 0, 5, 10, 20 and 30 years of missing data. Missing data are reported as the percentage of months where data are available to the total number of months, that is, $X = 100, 90, 76, 51$ or 25% . Each dot indicates a station, and each colour indicates the percentage P_x of data available at the station. Maps are shown for pressure levels at 925, 850, 500, 300, 100 and 30 hPa. In brackets, the total number of stations with a certain percentage of data is reported [Colour figure can be viewed at wileyonlinelibrary.com]

sampling (i.e., 100, 90, 76, 51 and 25%). Hereinafter, the percentage of available data at each pressure level will be indicated as P_x , that is, if we are referring to the number of stations with at least 90% of the data available at a certain pressure level for a specific test, this will be reported as P_{90} . In the following, “ da^{-1} ,” instead of “ decade^{-1} ,” is used to ease the reading of the text and results provided.

In Figure 1, the total number of stations recording large amounts of data (e.g., at least 90%) becomes extremely limited at the top of the atmosphere. At 925 hPa (Figure 1, top-left panel and Figure 2), no station among the total number of stations at the global scale reaches the full expected sampling (i.e., have full time

series without gaps). This is particularly because the 925 hPa pressure level was only adopted at the global scale by the end of 1991, although it has been planned since 1977 (WMO, 1977; Oakley, 1993).

No stations among the total number of stations ($n = 144$) at the Tropics have at least P_{63} of monthly data for all levels from 150 to 10 hPa owing to the paucity or complete lack of observations and to sensor limitations to produce suitable records (Figure 2, middle panel). In the SH ($n = 85$), the number of stations with P_{63} decrease substantially from 30 stations to less than 10 stations between 200 and 150 hPa (Figure 2, right panel). In both regions, there is a small number of stations with a substantive data record in the UT and LS, likely insufficient

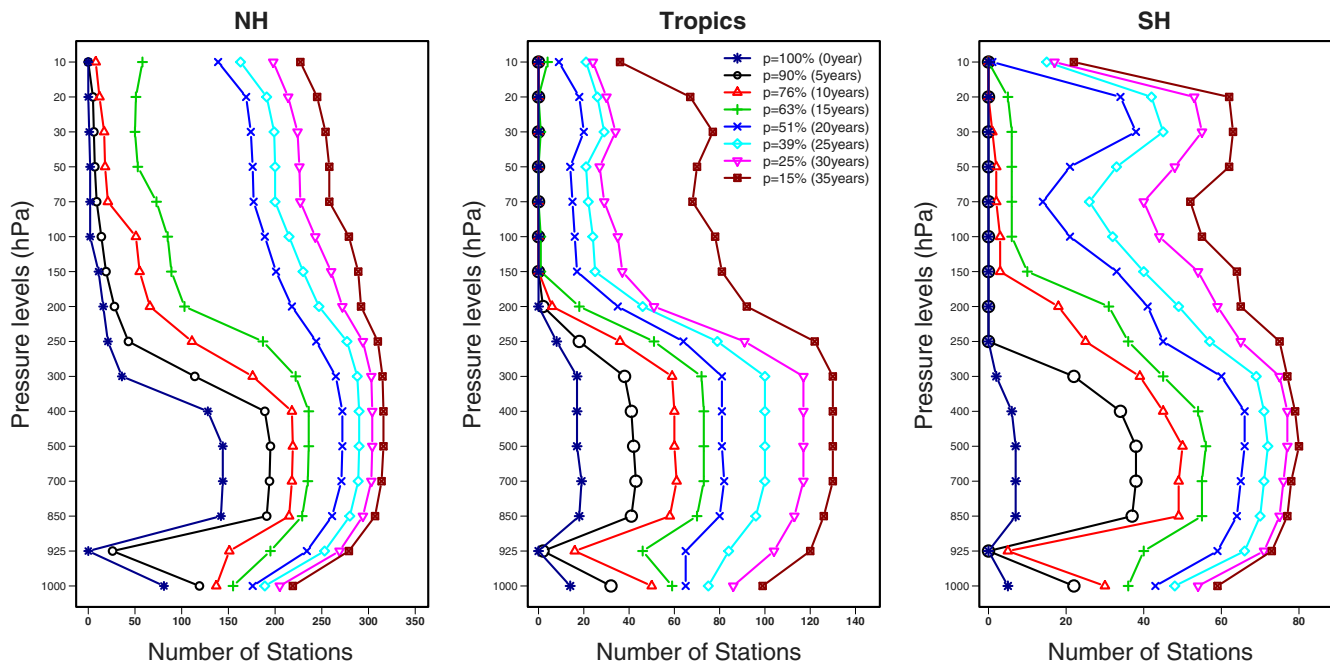


FIGURE 2 Number of radiosonde stations recording a given percentage P_x of temperature and relative humidity data at each pressure level since 1978–2018 for different latitudinal belts: that is, in the Northern Hemisphere (NH, $20^\circ\text{N} < \text{latitude} < 90^\circ\text{N}$), in the Tropics ($|\text{latitude}| < 20^\circ$) and in the Southern Hemisphere (SH, $90^\circ\text{S} < \text{latitude} < 20^\circ\text{S}$). In brackets, the amount of missing data in number of years is reported [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

to allow a reliable climate trend detection (Weatherhead *et al.*, 1998; Schlegel and Smit, 2016). Over the Northern Hemisphere (NH) ($n = 427$), where most of the radiosonde stations operate (70% of radiosonde launches are located between 30 and 60°N), the number of available observations for all P_x are more abundant (with a minimum of 10 stations reporting data at all pressure levels) (Figure 2). To evaluate the sensitivity of decadal trends to regression methods and to the effects of subsampling, the samples P_{100} , P_{90} , P_{76} and P_{51} are considered. Note that observations corresponding to P_{51} were considered to quantify the uncertainties for those applications based on a subsample of the IGRA dataset, including stations with a significant missing data. This extends several previous studies investigating trend differences solely considering one single percentage of data availability (Gaffen *et al.*, 2000).

3 | METHODOLOGY

3.1 | Statistical methods

Regression is a statistical method used to estimate the trend component of a climate data record (Mudelsee, 2019). A multitude of methods has been provided in the literature to estimate trends. Regression model building is broadly used

in literature to study the relationship between a dependent variable and a set of independent variables. Outliers from observation measurements in linear regression models are often encountered, and this is also true for radiosounding observations. Outliers can affect the variability associated with the heavy-tailed and skewed nature of data distributions but lead to a dramatic change of the magnitude of linear regression coefficients estimated and even the direction of coefficient signs (Choi, 2009; Wilks, 2011). At present, a reference dataset (Thorne *et al.*, 2017), other than GRUAN (The Global Climate Observing System [GCOS]-Reference Upper-Air Network; <https://www.gruan.org>), which comprises a small number of high-quality stations (15–20) with considerable data records since 2010, does not exist. Therefore, the analysis of trend estimations has been typically carried out on different datasets, selecting different station networks and periods of record using different regression methods. However, it appears still challenging to provide a “best” trend estimation for the different ECVs from radiosounding measurements due to the several reasons discussed earlier.

The calculation of trends takes into account time series autocorrelation (Weatherhead *et al.*, 1998; Weatherhead *et al.*, 2002), which is often present in climate data records (e.g., temperature). In this paper, the autocorrelation was taken into account by considering the first-order autoregressive model (Weatherhead

et al., 1998; Weatherhead *et al.*, 2002; Mudelsee, 2019) by assuming a simple linear regression model of the form:

$$y_t = \alpha + \beta x_t + \mu_t, t = 1, \dots, T \quad (1)$$

where y_t are the monthly anomaly time series, t the time variable assigned to y_t , α a constant term, x_t the linear trend function, β the linear trend and u_t the residual term that is assumed to be autoregressive of the order of 1 [AR (1)] when monthly data are considered (Weatherhead *et al.*, 2002; Hartmann *et al.*, 2013) and may be related as follow:

$$u_t = \rho u_{t-1} + \varepsilon_t, t = 2, \dots, T \quad (2)$$

where ρ determines the sign of this autocorrelation effect and is thus designated as the coefficient of autocorrelation. ε_t (the white noise series) are the independent random variables with mean zero and common variance σ_ε^2 :

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2), t = 1, \dots, T \quad (3)$$

This approach is similar to that one used in Intergovernmental Panel on Climate Change (IPCC) (Hartmann *et al.*, 2013). We also assumed that $-1 < \rho < 1$, so the noise process u_t is stationary. This model allows the noise to be autocorrelated among successive observations, such that $\rho = \text{Corr}(u_t, u_{t-1})$ (Weatherhead *et al.*, 1998; Mudelsee, 2019).

3.2 | Methods of estimating trends

Trend estimation methods applied to estimate linear trends are broad and fall generally into two main categories: parametric and nonparametric. The aim of this section is to extend several previous studies investigating trend estimation uncertainties, focusing on the difference between pairs of regression methods (most often a parametric vs. a nonparametric, e.g., Finger *et al.*, 1995; Gaffen, 1994; Gaffen *et al.*, 2000; Parker, 1985; Santer *et al.*, 2000). The sensitivities of regression methods are then assessed at each mandatory level for each P_x for the following methods:

1. Simple linear regression (hereinafter LIN) is probably the most commonly considered analysis method when examining the relationship between a quantitative outcome and a single quantitative explanatory variable. The term « simple » is used to clarify that only a single explanatory variable is considered. As discussed above, LIN is based on statistical significance via a

Student's t test and consists of a parametric technique not resistant to outliers (Lanzante, 1996) and has been demonstrated in many investigations to be less accurate and non-robust for skewed and heteroskedastic datasets (Lanzante, 1996; Wilks, 2011; Dervilis *et al.*, 2015).

2. LANZANTE (hereinafter LAN) is a resistant and non-parametric regression based on the median of pairwise slopes regression (Hoaglin *et al.*, 1986, 1893; Lanzante, 1996). It involves the computation of the slopes between every possible pair of points in the time series, taking the median value as the trend estimate. It was first used in the assessment of hydro-climatic trends (Lettenmaier *et al.*, 1994). It is called the Theil-Sen estimator and can be significantly more accurate than simple linear regression for skewed and heteroskedastic data and shows performances similar to non-robust least squares even for normally distributed data in terms of statistical power (Theil, 1950; Siegel and Benson, 1982; Helsel and Hirsch, 1992). The breakdown bound was estimated at about 29% by Theil (1950) and was extended to 50% by Siegel and Benson (1982), making the regression method very robust. In addition, if the errors are normally distributed, and no outliers are present, the estimators were found very similar to classic least squares (Lanzante, 1996).
3. Least absolute deviation (hereinafter LAD) is a resistant and nonparametric regression method fitting the paired data to the linear model using a robust and resistant LAD method (Rice and White, 1964; Barrodale, 1968; Wong and Schneider Jr, 1989; Calitz and R  ther, 1996; Santer *et al.*, 2000). The technique is based on an algorithm by Barrodale and Roberts (1974). Comparisons with the classical least-squares regression have been made to demonstrate that the proposed algorithm is more resistant to the existence of outliers and gives more intuitive results with less sensitivity to outliers (Wong and Schneider Jr, 1989). In addition, Calitz and R  ther (1996) found that the algorithm gives similar results to the classical least-squares regression in the absence of outliers. Wong and Schneider Jr (1989) reached similar conclusions and showed that the algorithm has no distributional or independence assumptions.
4. LMROB (hereinafter LMR) is a robust and nonparametric regression method based on an estimator for linear regression models (Finger, 2010; Koller and Stahel, 2011; Susanti *et al.*, 2014). The procedure proposed in the algorithm has been implemented in the R-package "robustbase". It provides different robust regression techniques, as well as robust univariate and multivariate methods (Rousseeuw and

Hubert, 2011). Apart from robust regression techniques for linear regression analysis, it includes robust methods for nonlinear and generalized linear models. It uses a bi-square redescending score function and returns a highly robust and efficient estimator with 50% breakdown point and 95% asymptotic efficiency for normal errors. The breakdown point is a measure of resistance to misbehaviour of the datasets, while efficiency measures the performance relative to some standard, such as a parametric estimator or a test assuming a Gaussian distribution. This method is strongly recommended for the robust estimation of linear regression due to its high efficiency when the errors have normal distribution (Finger, 2010).

The trend estimation methods used to estimate linear trends were selected from among the most frequently considered in climate studies using in-situ observations. Decadal trends are estimated from monthly anomaly time series at all selected stations and at all pressure levels. Resistant procedures are generally less efficient when the underlying distribution is Gaussian but provide much better results when data contain outliers or when the distribution is non-Gaussian.

3.3 | Subsampling strategies

Estimated subsampling effects may differ considerably among investigations and can be dependent on the approach adopted to subsample a global dataset (e.g., Rosen *et al.*, 2003; Free and Seidel, 2005; McCarthy, 2008). Referring to the Geophysical Fluid Dynamics Laboratory model output, Oort (1978) was among the first to quantify the sampling errors from global tropospheric upper-air temperatures due to spatial gaps. Results found that differences range from 0.50 to 1.0 K at individual pressure levels for an 855-station network and conclude that the network was generally sufficient for quantifying the statistics and trends of large-scale circulation in the NH but not in the SH. According to the Angell 63-station network (Angell and Korshover, 1975), Trenberth and Olson (1991) estimated subsampling errors using ECMWF operational analyses and found differences between the sampling effects estimated at the global scale in the NH and the SH. From global analyses of MSU and reanalysis datasets, Agudelo and Curry (2004) found that the difference between tropospheric trends in the global dataset and sampling dataset, according to (Lanzante *et al.*, 2003) the 87-station network, is up to $0.08 \text{ K}\cdot\text{da}^{-1}$. They concluded that the network overestimated temperature decadal trends due to the shortage of observations over the oceans.

Several studies have been focused on determining sampling effects estimated from different particular network configurations and sizes (e.g., Trenberth and Olson, 1991; Santer *et al.*, 2000; Agudelo and Curry, 2004; Free and Seidel, 2005; McCarthy, 2008). In addition, varying sampling density was tested to determine the theoretical size requirements for a network to provide sampling errors below a predefined threshold. McCarthy (2008) found that, to keep subsampling effects below $0.05 \text{ K}\cdot\text{da}^{-1}$ in the troposphere and $0.10 \text{ K}\cdot\text{da}^{-1}$ in the stratosphere, a radiosonde station at approximately every 30° longitude and 15° latitude north of 30°N is required. Results concluded that radiosonde network requirements are inadequate for monitoring humidity in the Tropics and SH due to the large inhomogeneous station distribution. In addition, the geographical coverage of radiosonde stations changed over time, and the periods of records vary among the stations and are characterized by several gaps. The effects of missing data in trend estimations have been investigated by Schlegel and Smit (2016). Results found that the percentage of missing data (% NA) in the time series of datasets and measurement accuracy are the variables that mainly affect accurate trend estimations.

The strategy adopted in this work to quantify spatial subsampling effects is to calculate the differences between decadal trends estimated for different subsets of radiosounding stations artificially selected versus the complete dataset. The different subsets of the stations are randomly selected for each P_x and each latitude to reduce the effects of the irregular spatial distribution of the stations (Figure 1) and their different time coverage at different pressure levels (Figure 2). Temporal subsampling effects or effects due to missing data are also quantified using the differences between decadal trends estimated for two different values of P_x (i.e., $P_{x1} - P_{x2}$). To analyse the statistical significance of the sensitivity of trend estimation to regression methods and to data subsampling, the rank-based nonparametric Wilcoxon-Mann-Whitney (WMW) test (Siegel and Benson, 1982) is used. The WMW test is used to test whether differences of estimated trends from two regression methods (or from two samples of datasets) are significantly different from zero with a 95% confidence interval. The main reason for using the nonparametric WMW statistical test is that it tends to be more powerful and better suited for non-normally distributed data compared to parametric tests such as Student's *t* test (Sy and Quesada, 2020).

4 | SENSITIVITY OF TRENDS TO THE CHOICE OF LINEAR REGRESSION METHODS

Figure 3 shows the comparison of the temperature decadal trend differences due to the use of different linear

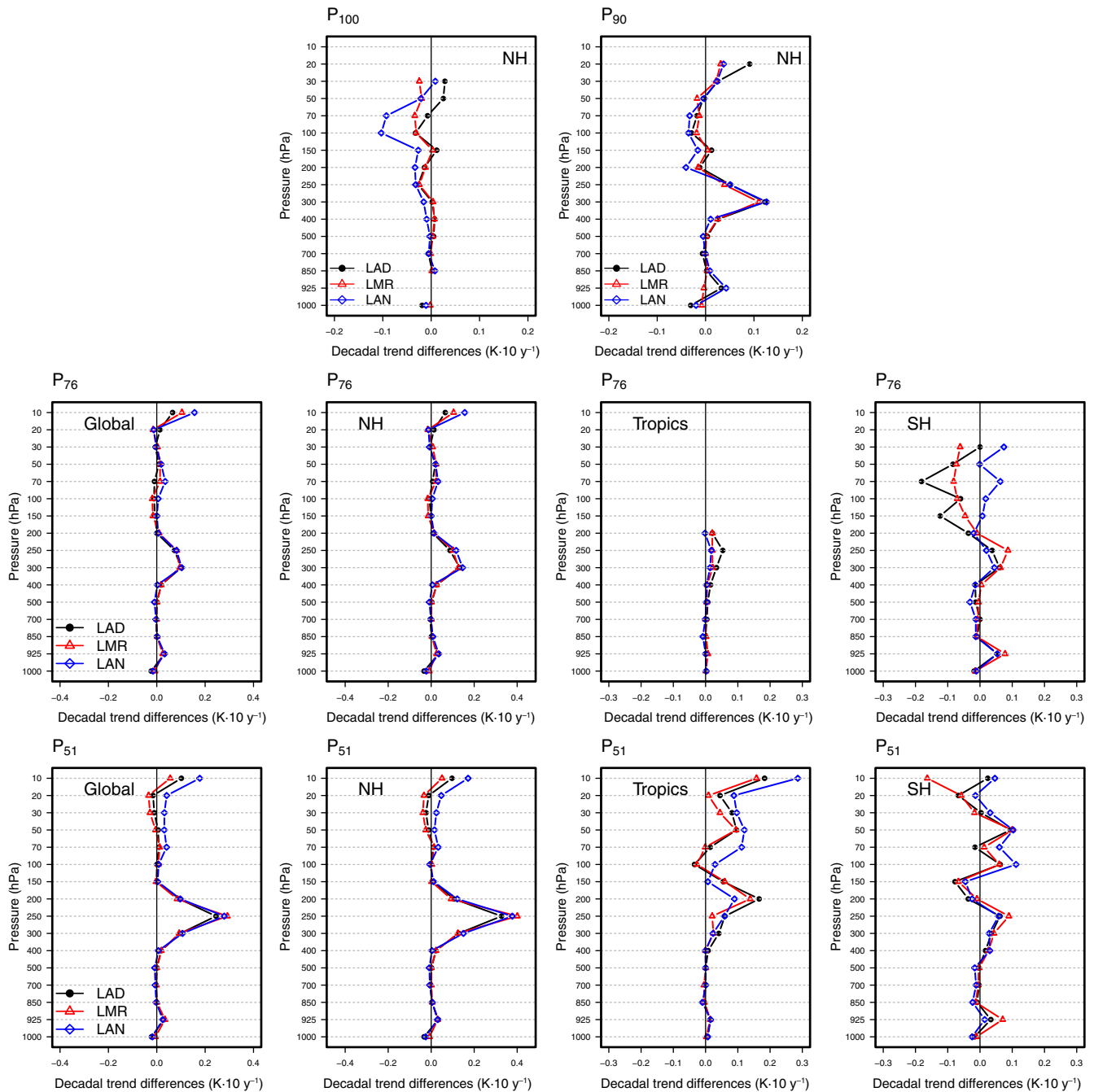


FIGURE 3 Temperature decadal trend differences (K decade^{-1}) estimated between pairs of nonparametric linear regression methods with respect to the simple parametric linear method (LIN), that is, (LAD – LIN; black curve), (LMR – LIN; red curve) and (LAN – LIN; blue curve) for each of the different pressure levels. Due to the limited observation available, differences calculated based on observations of P_{100} and P_{90} (top panels) are only shown for the Northern Hemisphere (NH, $20^\circ\text{N} < \text{latitude} < 90^\circ\text{N}$), while trend differences for P_{76} (middle panels) and for P_{51} (bottom panels) datasets are considered in all latitudes: that is, at global scale, in the Tropics ($|\text{latitude}| < 20^\circ$), in the Northern Hemisphere (NH, $20^\circ\text{N} < \text{latitude} < 90^\circ\text{N}$) and in the Southern Hemisphere (SH, $90^\circ\text{S} < \text{latitude} < 20^\circ\text{S}$) [Colour figure can be viewed at wileyonlinelibrary.com]

regression methods with respect to the LIN method, that is, (LAD – LIN; black curve), (LMR – LIN; red curve) and (LAN – LIN; blue curve) for the datasets corresponding to P_{100} , P_{90} , P_{76} and P_{51} , respectively. Due to the observation availability, P_{100} and P_{90} datasets are

only considered in the NH (Figures 1 and 2). Figure 4 shows the same as Figure 3 but for RH.

For both temperature and RH, the variability of the difference along the vertical profile for all the regression methods increases substantially from P_{100} to P_{51} .

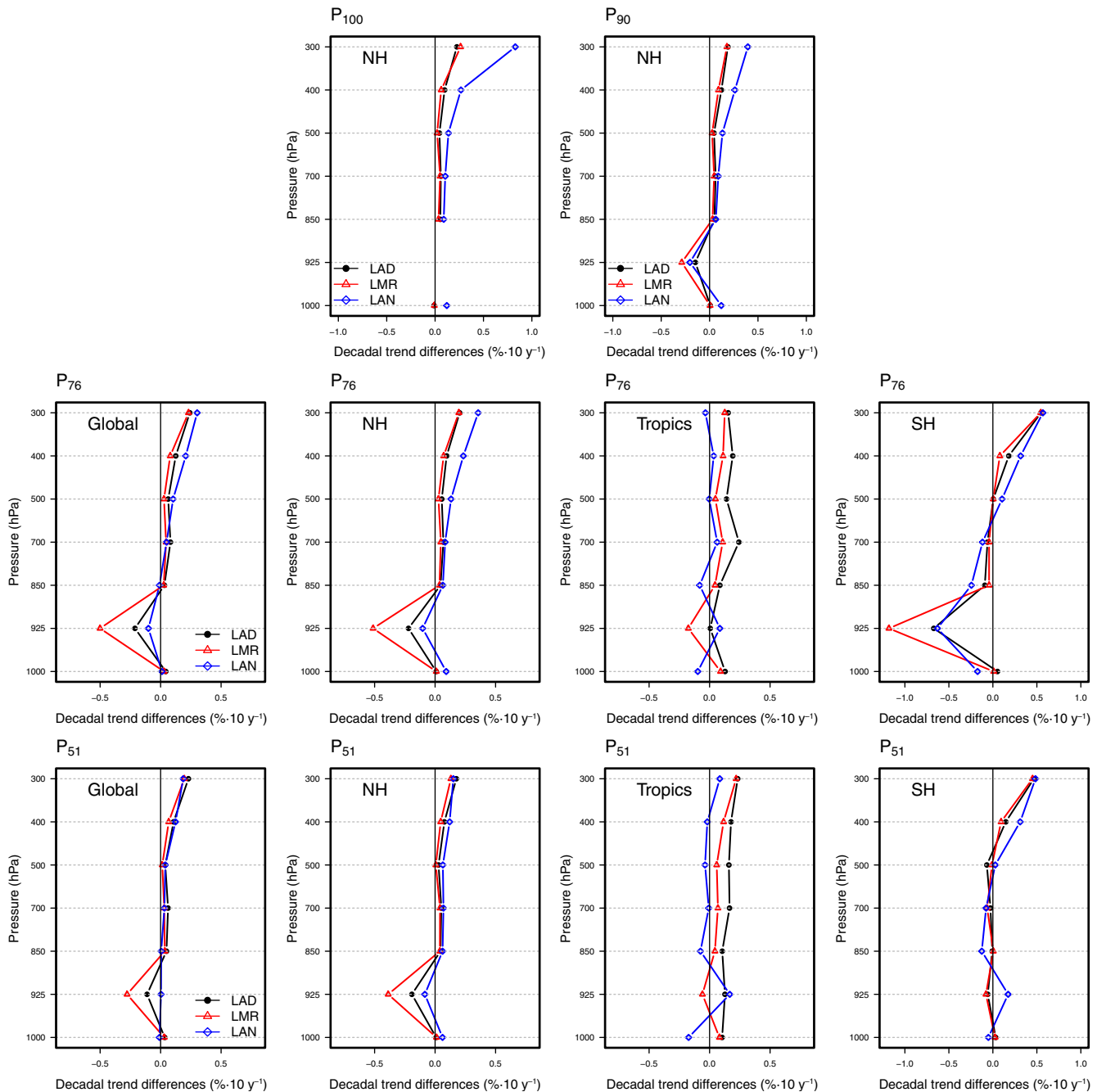


FIGURE 4 Same as Figure 3 but for relative humidity decadal trend differences ($\% \cdot \text{decade}^{-1}$) [Colour figure can be viewed at wileyonlinelibrary.com]

However, the comparison between the regression methods for temperature shows that decadal trends estimated with LIN significantly differ from the trends estimated with the three nonparametric methods, particularly between 400 and 200 hPa and at pressure levels below 100 hPa. Beyond the spatial and temporal sampling, the spread in the trends estimates obtained with different regression methods is influenced by several factors, such as: (a) sensitivity of the regression methods due to the specific nature and features of each dataset, that is,

variability, outliers and measurements uncertainty (e.g., Lanzante, 1996; Weatherhead *et al.*, 1998; Weatherhead *et al.*, 2002); (b) existence and timing of the largest change points, that is, structural breaks in the time series due to various reasons, such as instrumental effects or station re-locations (e.g., Gaffen *et al.*, 2000; Santer *et al.*, 2000); and (c) length, stability and number of missing data in the data records (e.g., Gaffen, 1994; Weatherhead *et al.*, 1998; Gaffen *et al.*, 2000; Santer *et al.*, 2000; Schlegel and Smit, 2016).

The P_{100} dataset shows that trend differences among regression methods vary from $-0.10 \text{ K}\cdot\text{da}^{-1}$ (LAN/LIN) to $-0.01 \text{ K}\cdot\text{da}^{-1}$ (at LAD/LIN and at LMR/LIN trend differences) below 100 hPa. The spread, compared to P_{90} , is larger and can be due to several aspects, such as selecting fewer stations, while with more complete data records, the non-normality and skewness of the distribution may increase the uncertainty for LMR and LAD methods (e.g., Wong and Schneider Jr, 1989; Finger, 2010). Conversely, if the datasets are normally distributed and homoscedastic, the LAN method may be affected by a large type I error (arriving at a false positive) and thus may be penalized with respect to other methods. Finally, if outliers and/or large change points are present in the datasets, both LMR and LAD regression methods may be more sensitive (i.e., less resistant) than LAN and thus exhibit small differences with respect to the LIN method. Nevertheless, the values of the trends for the P_{100} dataset give a clear idea of the uncertainty outcome from the trend estimation due to the use of one regression method or another for the considered dataset.

Similar results are obtained in the NH for both P_{90} and P_{76} datasets, with maximum differences around $0.15 \text{ K}\cdot\text{da}^{-1}$ at 300 hPa. Considering P_{76} , at the global scale, differences among regression methods may also exceed $0.15 \text{ K}\cdot\text{da}^{-1}$ at 300 hPa, despite the large number of observations available. At the Tropics, regression methods give similar results at all pressure levels with a maximum difference of around $0.05 \text{ K}\cdot\text{da}^{-1}$ at 250 hPa. In SH, the differences between regression methods are around $0.10 \text{ K}\cdot\text{da}^{-1}$ at 925 and 250 hPa, with large discrepancies ranging from $-0.20 \text{ K}\cdot\text{da}^{-1}$ (LAD/LIN) to $0.08 \text{ K}\cdot\text{da}^{-1}$ (LAN/LIN) at 70 hPa, mainly due to limited amount of observations and the consequent limited spatial representativeness of the dataset.

Similar results for P_{51} are shown in the bottom panel of Figure 3: Differences among the regression methods become significant at all latitudes compared to P_{100} , P_{90} and P_{76} , with trend differences exceeding $0.40 \text{ K}\cdot\text{da}^{-1}$ at 250 hPa in the NH. In the Tropics, at pressure levels smaller than 300 hPa, where observations become limited, the spread among regression methods varies from $0.20 \text{ K}\cdot\text{da}^{-1}$ (LAD/LIN) to $0.10 \text{ K}\cdot\text{da}^{-1}$ (LAN/LIN) at 200 hPa. In SH, large trend differences are also estimated at pressure levels below 300 hPa, with spread among methods varying from $-0.15 \text{ K}\cdot\text{da}^{-1}$ (LMR/LIN) to $0.05 \text{ K}\cdot\text{da}^{-1}$ (LAD/LIN).

Comparing trend differences estimated from P_{76} and P_{51} at the global scale and over the NH (Figure 3 middle and bottom panels), results show that differences among regression methods estimated from P_{51} are four times larger than those estimated from P_{76} at 250 hPa. Comparing also trend differences for both percentage of datasets

P_{76} and P_{51} (Figure 3) in the Tropics at 200 hPa, differences estimated from P_{51} datasets are at least two times larger than differences estimated from P_{76} . This means that differences among the regression methods maybe at least two times smaller if the coverage of available observations increased from 51 to 76% of the total sampling. This is consistent with the fact that trends calculated using data records with significant temporal gaps of missing data (>20 years) are typically affected by large sampling uncertainties. In other words, missing data values increase the possibility to detect a trend in a stationary time series (type-2 error, i.e., finding false negatives), also increasing the discrepancies among different regression methods.

On the other hand, adding more stations with large temporal gaps to the stations with smaller gaps (e.g., from P_{76} to P_{51}) may substantially increase the spreads among regression methods and, therefore, induce larger uncertainties in trend estimations. In order to reduce uncertainties derived from the choice of an estimation method, the stations with the most complete data records must be selected. This is evident for the trends estimated at the global scale and in the NH at pressure levels lower than 300 hPa in the UTLS regions, where measurement coverage is more limited than at other atmospheric levels.

Table 1 summarizes the results shown in Figure 2 at four atmospheric layers: lower troposphere, middle troposphere, upper troposphere and lower stratosphere (hereinafter, LT, MT, UT and LS, respectively). The averaged differences among the regression methods are smaller for P_{100} and P_{90} (i.e., less than $-0.02 \text{ K}\cdot\text{da}^{-1}$ for P_{100} and around $0.06 \text{ K}\cdot\text{da}^{-1}$ for P_{90} in the UT) but statistically significant. Differences also range around $\pm 0.05 \text{ K}\cdot\text{da}^{-1}$ at all latitudes except in SH, where differences become statistically significant and can reach up to $-0.10 \text{ K}\cdot\text{da}^{-1}$ in the LS for P_{76} . From P_{51} , differences are statistically significant as well and can reach up to $0.15 \text{ K}\cdot\text{da}^{-1}$ in the UT, likely related to the large variability of the dynamical processes governing stratosphere-troposphere exchange, as well as to the paucity of radio-sonde records in the UTLS region.

In Figure 4 and Table 2, a similar comparison as in Figure 3 and Table 1, respectively, is reported but for RH. Comparing trend differences at 300 hPa for all P_x , a large spread among regression methods from $0.20\%\cdot\text{da}^{-1}$ (LMR/LIN) to $0.80\%\cdot\text{da}^{-1}$ (LAN/LIN) is obtained for P_{100} . A large spread at 925 hPa is shown in Figure 4 for all P_x . For P_{76} , the trend differences range within $-0.50\%\cdot\text{da}^{-1}$ (LMR/LIN) and $-0.05\%\cdot\text{da}^{-1}$ (LAN/LIN) in the NH, while in the SH, the differences are larger and range from less than $-1.0\%\cdot\text{da}^{-1}$ (LMR/LIN) to $-0.50\%\cdot\text{da}^{-1}$ (LAD/LIN). The large differences at 300 hPa and 925 hPa

TABLE 1 Temperature decadal trend differences ($K \cdot decade^{-1}$) estimated between pairs of nonparametric linear regression methods with respect to the simple parametric linear method (LIN), that is, (LAD – LIN), (LMR – LIN) and (LAN – LIN), for each of the four atmospheric layers: lower stratosphere (LS), upper troposphere (UT), middle troposphere (MT) and lower troposphere (LT)

Temperature decadal trends ($K \cdot 10 y^{-1}$)									
NH									
Atmospheric layers									
P_{100}					P_{90}				
	LAD – LIN	LMR – LIN	LAN – LIN	LAD – LIN	LMR – LIN	LAN – LIN	LAD – LIN	LMR – LIN	LAN – LIN
LS	0.012	-0.026	-0.043	-0.001	0.071	0.039	-0.004	0.062	0.039
UT	-0.007	0.004	-0.006	-0.004	0.000	0.000	0.000	0.000	0.000
MT	0.005	0.000	-0.006	-0.001	0.000	0.000	0.000	0.000	0.000
LT	-0.002	0.000	-0.001	-0.001	0.000	0.000	0.000	0.000	0.000
Global					SH				
Atmospheric layers					Tropics				
	LAD – LIN	LMR – LIN	LAN – LIN	LAD – LIN	LMR – LIN	LAN – LIN	LAD – LIN	LMR – LIN	LAN – LIN
LS	-0.002	0.018	0.066	0.039	0.021	0.014	0.006	0.018	0.016
UT	0.060	0.060	0.066	0.039	0.021	0.014	0.072	0.071	0.086
MT	0.036	0.034	0.028	0.018	0.010	0.006	0.048	0.045	0.041
LT	0.000	0.002	0.000	0.002	0.001	-0.001	0.000	0.004	0.001
P_{76}					P_{51}				
LS	-0.002	0.010	0.018	0.006	0.018	0.016	-0.106	-0.074	0.039
UT	0.060	0.060	0.066	0.039	0.021	0.014	0.072	0.071	0.086
MT	0.036	0.034	0.028	0.018	0.010	0.006	0.048	0.045	0.041
LT	0.000	0.002	0.000	0.002	0.001	-0.001	0.000	0.004	0.001
P_{100}					P_{90}				
LS	0.000	-0.007	0.034	-0.008	0.024	0.025	-0.016	0.022	0.057
UT	0.100	0.108	0.113	0.062	0.040	0.044	0.123	0.136	0.026
MT	0.038	0.036	0.034	0.016	0.008	0.007	0.049	0.048	0.015
LT	0.001	0.005	-0.001	0.002	0.001	0.001	0.001	0.006	0.000

Note: Differences are calculated based on radiosounding observations corresponding to P_{100} , P_{90} , P_{76} and P_{51} , respectively, at global scale, in the Tropics ($|\text{latitude}| < 20^\circ$), in the Northern Hemisphere (NH, $20^\circ N < \text{latitude} < 90^\circ N$) and in the Southern Hemisphere (SH, $90^\circ S < \text{latitude} < 20^\circ S$). Due to the limited number of observation available, P_{100} and P_{90} datasets are considered in the NH only. Decadal trend differences statistically significant with 95% confidence intervals, that is, differences that passed the WMW test at 0.05 level, are reported in bold.

TABLE 2 Same as Table 1 but for relative humidity decadal trend differences (%.decade⁻¹) in the middle troposphere (MT) and lower troposphere (LT)

Relative humidity decadal trends (%.10 y ⁻¹)													
NH													
Atmospheric layers	P ₁₀₀			P ₉₀									
	LAD - LIN	LMR - LIN	LAN - LIN	LAD - LIN	LMR - LIN	LAN - LIN	LMR - LIN	LAN - LIN	LMR - LIN	LAN - LIN			
MT	0.066	0.088	0.272	0.107	0.086	0.241							
LT	0.042	0.029	0.112	0.038	0.018	0.089							
SH													
Atmospheric layers	Global			NH									
	LAD - LIN	LMR - LIN	LAN - LIN	LAD - LIN	LMR - LIN	LAN - LIN	LAD - LIN	LMR - LIN	LAN - LIN	LMR - LIN			
	0.137	0.105	0.198	0.161	0.094	0.094	-0.000	0.111	0.091	0.232	0.221	0.186	0.310
	0.022	-0.034	0.023	0.141	0.056	0.056	-0.018	0.005	-0.055	0.061	-0.049	-0.052	-0.113
P ₇₆													
	0.124	0.091	0.115	0.189	0.131	0.131	0.009	0.092	0.060	0.111	0.172	0.169	0.267
	0.016	-0.031	0.016	0.133	0.041	0.041	-0.026	-0.008	-0.054	0.035	-0.029	-0.027	-0.010

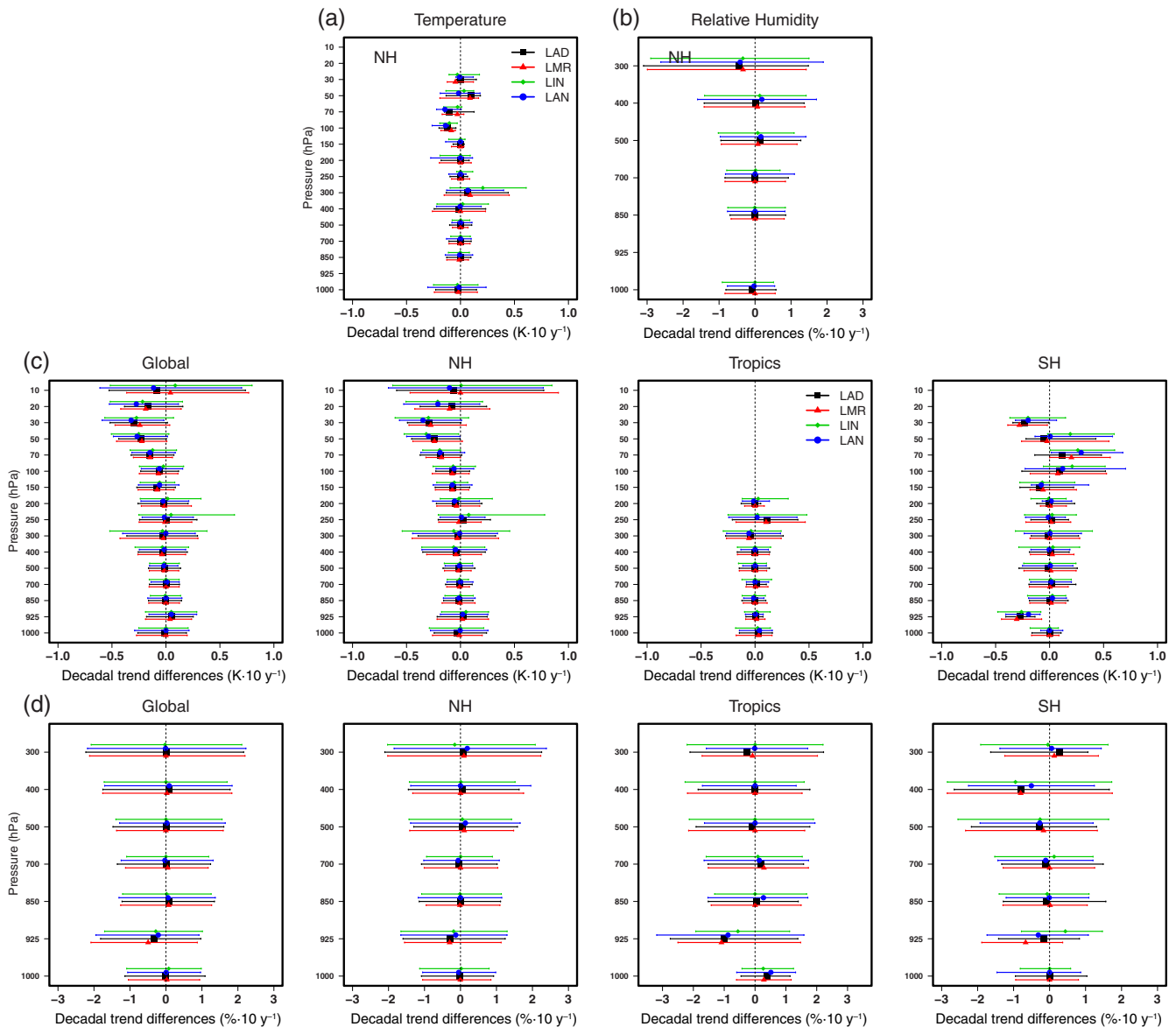


FIGURE 5 Sensitivity of temperature ($\text{K}\cdot\text{decade}^{-1}$) and relative humidity ($\%\cdot\text{decade}^{-1}$) decadal trends to the missing data effects. The effect is quantified using the differences between decadal trends estimated for two different values of P_x (i.e., $P_{x1} - P_{x2}$). Due to the limited observation available, trend differences calculated between P_{100} and P_{90} datasets (i.e., $P_{100} - P_{90}$) are only shown for the Northern Hemisphere (NH, $20^\circ\text{N} < \text{latitude} < 90^\circ\text{N}$) (top panels (a) for temperature and (b) for relative humidity datasets). Differences between decadal trends estimated for P_{76} and for P_{51} (i.e., $P_{76} - P_{51}$) for temperature ((c), middle panels) and for relative humidity ((d), bottom panels) are shown at the global scale and at all the latitudinal belts: Tropics ($|\text{latitude}| < 20^\circ$), Northern Hemisphere (NH, $20^\circ\text{N} < \text{latitude} < 90^\circ\text{N}$) and Southern Hemisphere (SH, $90^\circ\text{S} < \text{latitude} < 20^\circ\text{S}$). Dots are representative of the median values for each of the different regression methods (different colours), while horizontal bars are representative of the first and third quartiles of the corresponding probability distribution [Colour figure can be viewed at wileyonlinelibrary.com]

for RH can be also due to the large water vapour variability in the boundary layer in LT and also to the restricted number of observations (McCarthy, 2008; Ladstädter *et al.*, 2011). Results provided in Table 2 further show that trend differences in the LT and MT are statistically significant, with values ranging within 0.06 and $0.20\%\cdot\text{da}^{-1}$ in the MT and within $0.01\%\cdot\text{da}^{-1}$ and $0.10\%\cdot\text{da}^{-1}$ in

the LT for both P_{100} and P_{90} datasets. For P_{51} , the maximum spread occurs in the MT, with significant differences exceeding $0.30\%\cdot\text{da}^{-1}$ in the SH, while in the Tropics, differences are generally smaller and non-statistically significant. The large spread among regression methods observed at all latitudes at 925 hPa, shown in Figures 3 and 4, and at 250 hPa for temperature

datasets can be mainly explained by the small amount of radiosonde data available, as mentioned above.

5 | SENSITIVITY OF TRENDS TO TEMPORAL SUBSAMPLING EFFECTS

In this section, we quantify the sensitivity of decadal trends to temporal subsampling effects (i.e., missing data or changes in the sampling frequency). The effects are estimated for temperature and for RH time series in the NH using differences between trends for P_{100} and for P_{90} at all pressure levels (Figure 5a,b, top panels). The effects are also estimated at all latitudinal belts using the difference between P_{76} and P_{51} for temperature (Figure 5c, middle panels) and for RH (Figure 5d, bottom panels). In Figure 5, dots are representative of the median values for each regression method (different colours), while horizontal bars are representative of the first and third quartiles of the corresponding probability distribution.

Based on trend differences estimated between P_{100} and P_{90} for temperature (Figure 5a), it can be inferred that the temporal subsampling effects are larger at 300 hPa and at pressure levels below 150 hPa, where there is a substantive increase in the sampling frequency. At 300 hPa, the trend differences due to the missing data and the high local variability can exceed $0.20 \text{ K}\cdot\text{da}^{-1}$ for LIN, while for the nonparametric regression methods, the differences are less than $0.10 \text{ K}\cdot\text{da}^{-1}$. At pressure levels below 150 hPa, the spread among regression methods becomes larger, with differences ranging within $\pm 0.10 \text{ K}\cdot\text{da}^{-1}$. From Figure 5b, it is also possible to provide similar results as Figure 5a but for RH. Results show large temporal subsampling effects exceeding $0.50\%\cdot\text{da}^{-1}$ at 300 hPa, mainly due to the small number of available observations and the large atmospheric variability (McCarthy, 2008; Ladstädter *et al.*, 2011).

The effects of missing data are also obtained using P_{76} and P_{51} . Figure 5c show statistically significant discrepancies among the regression methods at 925 hPa and at pressure levels below 200 hPa (with values exceeding $0.40 \text{ K}\cdot\text{da}^{-1}$ at 30 hPa), which are mainly due to the substantial increase of the missing data in the time series. These effects are also observed for RH (Figure 5d), with values exceeding $1.0\%\cdot\text{da}^{-1}$ at 925 hPa in the Tropics. For temperature in the SH, a large spread is observed at 925 hPa and at atmospheric levels below 150 hPa and at 925 and 400 hPa for RH, mainly linked to the sparseness of radiosounding stations and small amount of available data.

Table 3 further shows how the temporal subsampling effect at different atmospheric layers is generally small

for all regression methods (around $-0.01 \text{ K}\cdot\text{da}^{-1}$ in the LT, $0.05 \text{ K}\cdot\text{da}^{-1}$ in the MT and in the UT and within $\pm 0.02 \text{ K}\cdot\text{da}^{-1}$ in the LS), when the amount of missing data is limited (i.e., $P_{100} - P_{90}$). Also for RH, the averaged trend differences due to missing data are generally small (around $-0.02\%\cdot\text{da}^{-1}$ in the LT and $0.10\%\cdot\text{da}^{-1}$ in the MT). Similar results are obtained using P_{76} and P_{51} (Table 3, bottom rows), but in the NH and at the global scale. The effects of missing data are statistically significant in the LS, with trend differences exceeding $-0.20 \text{ K}\cdot\text{da}^{-1}$. Table 3 (bottom rows) reports analogous results for RH. Results show that, except for SH where uncertainties are larger (more than $0.30\%\cdot\text{da}^{-1}$ in MT and $0.40\%\cdot\text{da}^{-1}$ in LT), temporal subsampling effects are generally small (less than $-0.12\%\cdot\text{da}^{-1}$ in the LT and no more than $\pm 0.11\%\cdot\text{da}^{-1}$ in the MT). These uncertainties are mainly due to spatial and temporal inhomogeneities in the station distribution.

In summary, the analysis reported in this section allows users to quantify the temporal and spatial subsampling uncertainties for their own climate application based on radiosounding data. To reduce the uncertainty on a trend estimation due to missing data below $\pm 0.15 \text{ K}\cdot\text{da}^{-1}$ for temperature, selecting a dataset with at least 76% of data coverage (i.e., time series observations with at least 30-years of data coverage) is recommended, while for RH, no clear correlation can be found between the estimated uncertainties and the missing data effects.

6 | SENSITIVITY OF TRENDS TO SPATIAL SUBSAMPLING EFFECTS

Figure 6 shows the sensitivity of temperature decadal trends to the effects of spatial subsampling, that is, uncertainty on trend estimations due to the selection of different subsets of stations. Due to the limited observation availability and spatial coverage of the dataset in the Tropics and SH, the effects of spatial subsampling are quantified over the NH only. Spatial subsampling effects are estimated using the difference between decadal trends calculated for a subset of radiosounding stations artificially selected (from 20 to 100 stations) versus the decadal trends obtained for the complete dataset. In Figure 6, the median values for each regression method are reported, along with the first and third quartiles (horizontal bars).

At all pressure levels below 400 hPa, the spatial subsampling effect becomes relevant for all P_x , with a large spread among the regression methods. It becomes obvious that subsampling effects are larger when a small number of stations is considered (<40 stations). Therefore, when increasing the number of stations, a

TABLE 3 Sensitivity of temperature and relative humidity decadal trends to the missing data estimated at four atmospheric layers: lower stratosphere (LS), upper troposphere (UT), middle troposphere (MT) and lower troposphere (LT)

$P_{100} - P_{90}$																
NH																
Temperature decadal trends ($K \cdot 10 y^{-1}$)						Relative humidity decadal trends ($\% \cdot 10 y^{-1}$)										
Atmospheric layers	LIN	LAD	LMR	LAN	LIN	LAD	LMR	LAN	LIN	LAD	LMR	LAN				
LS	0.006	0.020	-0.014	-0.029												
UT	0.145	0.066	0.076	0.052												
MT	0.108	0.049	0.056	0.041	0.136	0.116	0.117	0.166								
LT	-0.011	-0.009	-0.010	-0.011	-0.034	-0.030	-0.023	-0.011								
$P_{76} - P_{51}$																
Atmospheric layers	Global				Tropics				NH				SH			
Temperature decadal trends ($K \cdot 10 y^{-1}$)																
	LIN	LAD	LMR	LAN	LIN	LAD	LMR	LAN	LIN	LAD	LMR	LAN	LAN			
LS	-0.181	-0.184	-0.163	-0.197					-0.215	-0.199	-0.180	-0.222	0.137	0.041	0.119	
UT	0.040	0.001	-0.007	-0.005	0.060	0.037	0.041	0.030	0.024	-0.027	-0.040	-0.032	0.073	0.093	0.075	
MT	0.055	0.001	-0.011	-0.011	0.035	0.032	0.036	0.021	0.057	-0.018	-0.040	-0.033	0.057	0.052	0.035	
LT	-0.006	-0.008	-0.009	-0.006	-0.007	-0.007	-0.006	-0.010	-0.015	-0.017	-0.017	-0.013	-0.008	-0.021	-0.012	
Relative humidity decadal trends ($\% \cdot 10 y^{-1}$)																
MT	-0.005	0.007	0.008	0.077	-0.019	-0.047	-0.056	-0.029	0.080	0.101	0.113	0.203	-0.334	-0.285	-0.317	-0.291
LT	-0.122	-0.116	-0.126	-0.114	-0.115	-0.107	-0.100	-0.107	-0.041	-0.027	-0.042	-0.015	-0.430	-0.449	-0.455	-0.533

Note: Differences are calculated for each of the different regression methods between trends estimated for P100 and for P90 datasets (P100-P90, top rows) in the Northern Hemisphere (NH, $20^{\circ}N < \text{latitude} < 90^{\circ}N$). Differences are also estimated between trends estimated for P76 and for P51 datasets (P76-P51) at global scale, in the Tropics ($|\text{latitude}| < 20^{\circ}$), in the Northern Hemisphere (NH, $20^{\circ}N < \text{latitude} < 90^{\circ}N$) and in the Southern Hemisphere (SH, $90^{\circ}S < \text{latitude} < 20^{\circ}S$). Decadal trend differences statistically significant with 95% confidence intervals, that is, differences that passed the WMW test at 0.05 level, are reported in bold.

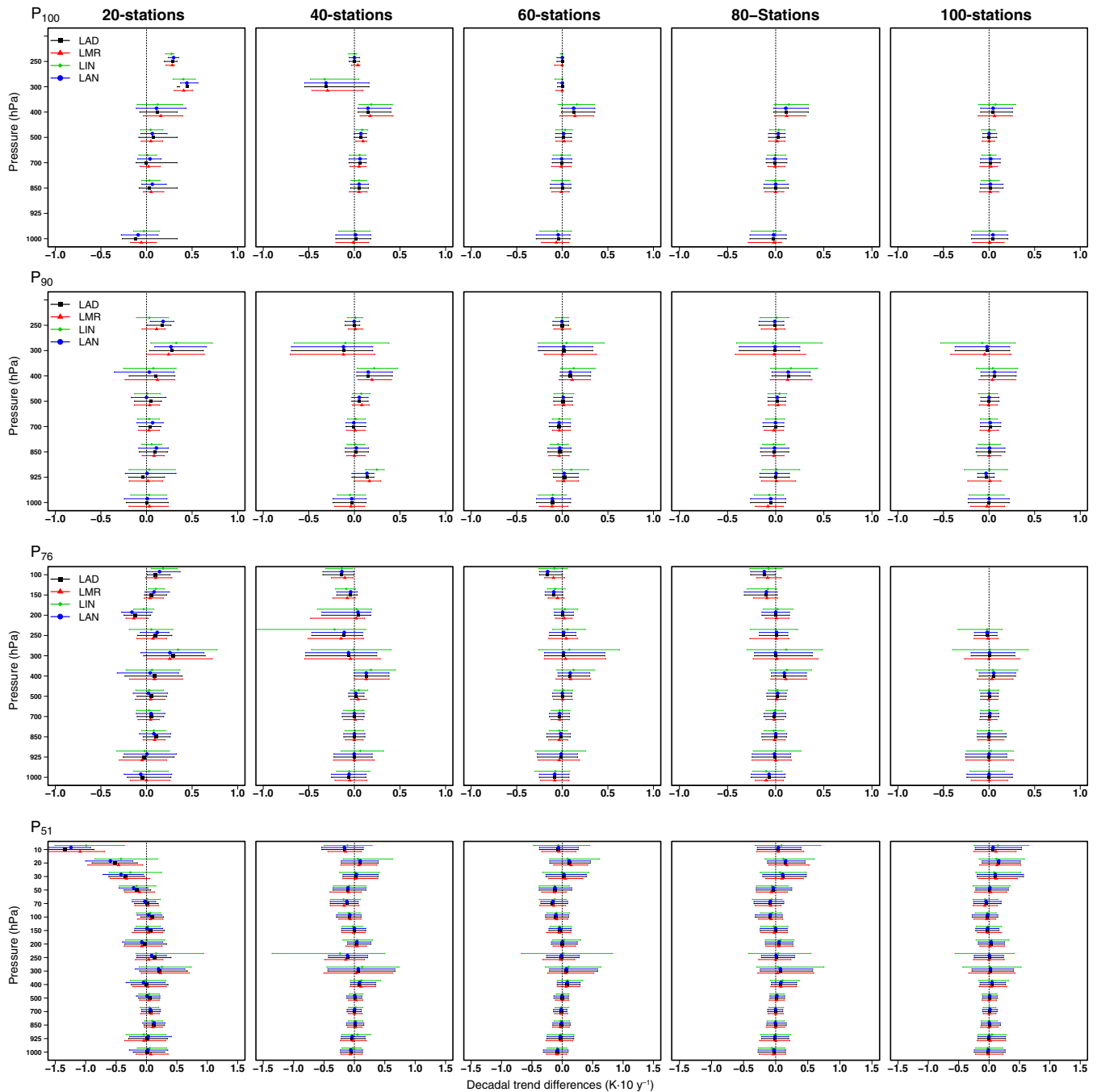


FIGURE 6 Sensitivity of temperature decadal trends ($\text{K}\cdot\text{decade}^{-1}$) to the spatial subsampling effects. The effects are estimated as the difference between trends estimated from each subset of radiosounding stations artificially selected (ranging from 20 to 100) versus the complete network in the Northern Hemisphere (NH, $20^\circ\text{N} < \text{latitude} < 90^\circ\text{N}$) only (due to the limited observation availability) at all mandatory pressure levels for P_{100} (top panels), for P_{90} (top-middle panels), for P_{76} (middle-bottom panels) and for P_{51} (bottom panels). The dots are representative of the median values for each of different regression methods (see different colours), while horizontal bars are representative of the first and third quartiles of the corresponding probability distribution [Colour figure can be viewed at wileyonlinelibrary.com]

convergence process is observed, with trend differences progressively decreasing to small values. This is true for P_{100} , P_{90} and P_{76} datasets, as shown in Figure 6. However, an increasing the number of stations (e.g., from 20 to

more than 100 stations), coupled with a limited amount of missing data, allows us to further reduce the noise coming from an inappropriate spatial and temporal sampling, thus reducing the uncertainties due to the choice

of a regression method. Above 400 hPa in the LT, where a dense distribution of observations is available, the subsampling effect becomes smaller, with differences generally less than $\pm 0.01 \text{ K}\cdot\text{da}^{-1}$, while differences increase when fewer than 20 stations are considered (exceeding $0.10 \text{ K}\cdot\text{da}^{-1}$ at 850 hPa).

Results provided in Table 4 also show that the spatial subsampling effects are generally small (lower than $\pm 0.10 \text{ K}\cdot\text{da}^{-1}$ in the LT and no more than $0.20 \text{ K}\cdot\text{da}^{-1}$ in the MT for all P_{100} and P_{90} datasets), but it is statistically significant, especially when trends are calculated from a moderate number of stations (e.g., less than 80). From P_{76} and P_{51} datasets, the effects of the spatial subsampling effects are also quite small and therefore negligible, with values ranging around $\pm 0.10 \text{ K}\cdot\text{da}^{-1}$. If fewer than 20 stations are considered, the differences become statistically significant, with values exceeding $0.2 \text{ K}\cdot\text{da}^{-1}$ for both the MT and the UT levels.

In Figure 7, similar results as Figure 6 are shown for RH. The spatial subsampling relative effects estimated are generally large at 1000 hPa and at 300 hPa, with values ranging within $\pm 0.50\%\cdot\text{da}^{-1}$ for all P_x and the number of stations considered. The calculated differences are mainly due to both the high variability of the humidity field in the LT and UT and to the biases and individual data processing typical of different radiosonde types. The most significant effects are observed below 400 hPa, with significant trend differences exceeding $-1.0\%\cdot\text{da}^{-1}$ for all P_x when 20 stations are considered, while they decrease asymptotically from 20 to 100 stations. Table 5 shows the relative effect of spatial subsampling in the LT, which is generally limited with values ranging around $\pm 0.10\%\cdot\text{da}^{-1}$ for almost all P_x , independent of the number of selected stations. Instead, in the MT, the differences become large and statistically significant for almost all datasets, especially for P_{51} , when the number of stations considered is fewer than 80: In this case, differences can be up to $-1.0\%\cdot\text{da}^{-1}$ when 20 stations are selected.

7 | DISCUSSION AND CONCLUSION

An extension of the analysis available in literature for the quantification of uncertainties in the estimation of decadal trends is presented in this paper. The effects of different regression methods, of missing data and of subsampling have been studied using historical radiosounding time series of temperature and RH from 656 stations in the period 1978–2018, available in version 2 of IGRA (Durre *et al.*, 2018). The strongest finding of this paper is that accurate estimations of long-term decadal trends in radiosounding time series may primarily

concern the choice of the regression method. When trends are calculated on the entire IGRA dataset (i.e., P_{100}), differences among regression methods may vary from $-0.10 \text{ K}\cdot\text{da}^{-1}$ to $-0.01 \text{ K}\cdot\text{da}^{-1}$ below 100 hPa for temperature and from 0.2 to $0.8\%\cdot\text{da}^{-1}$ at 300 hPa for RH. Differences can also increase up to $0.4 \text{ K}\cdot\text{da}^{-1}$ at 300 hPa when the amount of missing data exceeds 50% of the original dataset for temperature, while for RH, significant differences are observed in the LT at 925 hPa for almost all datasets. Such findings are consistent with previous studies available in the literature (e.g., Gaffen *et al.*, 2000; Santer *et al.*, 2000), but they are more complete because they are often limited to considering differences between pairs of regression methods and/or considering one percentage of data availability from datasets characterized by trend rates slower than those observed over the last two decades. Another important aspect of this paper is that this is, to our knowledge, the most recent work based on trends calculated using radiosounding datasets, which are now measuring using improved sensors in the epoch of the “climate change,” and therefore, it comes at the right moment as an improvement of past efforts.

In the past, analysis of uncertainties in trend estimations has been typically carried out on different datasets, selecting different station networks, periods of record and regression methods. From the different samples of dataset considered, as mentioned above, the spread among regression methods is influenced by several factors, such as: (a) sensitivity of the regression methods due to the specific nature and features of each dataset (Lanzante, 1996; Weatherhead *et al.*, 1998; Weatherhead *et al.*, 2002); (b) existence and timing of the largest change points (Gaffen *et al.*, 2000; Santer *et al.*, 2000); and (c) length, stability and amount of missing data in the data records (Schlegel and Smit, 2016). Further studies based on a comparison of several trend estimation methods using a larger number of stations with long data records are still a valid solution to improve our knowledge of climate trends and to quantify the related uncertainties.

Finally, two additional caveats are worth noting. First, the uncertainty contributions due to the effect of spatial subsampling biases and, second, the sensitivity of decadal trends to temporal subsampling effects (i.e., missing data). For instance, as we mentioned above, missing data are an intrinsic dataset characteristic due to a lack of continuity in operations contingently occurring at each measuring station. Although the completeness of historical radiosounding observations is improving over the years (Ferreira *et al.*, 2019), missing data in the time series are still frequent. This implies that, to provide a reliable assessment of the uncertainties affecting the

TABLE 4 Sensitivity of temperature decadal trends ($K \cdot decade^{-1}$) to subsampling effects at four atmospheric layers: lower stratosphere (LS), upper troposphere (UT), middle troposphere (MT) and lower troposphere (LT)

N°	20 stations			40 stations			60 stations			80 stations			100 stations									
	Atmospheric layers			Atmospheric layers			Atmospheric layers			Atmospheric layers			Atmospheric layers									
	LIN	LAD	LMR	LIN	LAD	LMR	LIN	LAD	LMR	LIN	LAD	LMR	LIN	LAD	LMR	LIN	LAD	LMR	LIN	LAD	LMR	
P ₁₀₀																						
LT	0.019	0.020	0.029	0.041	0.047	0.049	0.043	0.042	-0.016	-0.020	-0.018	-0.016	-0.013	-0.016	-0.016	-0.012	-0.001	0.006	-0.000	0.012		
MT	0.217	0.231	0.235	0.217	0.121	0.136	0.126	0.116	0.105	0.100	0.099	0.108	0.113	0.106	0.105	0.116	0.072	0.072	0.069	0.084		
P ₉₀																						
LT	0.032	0.038	0.037	0.035	0.022	0.019	0.016	0.010	-0.029	-0.034	-0.032	-0.034	-0.021	-0.024	-0.025	-0.021	-0.009	-0.001	-0.008	0.005		
MT	0.166	0.165	0.161	0.137	0.066	0.041	0.028	0.014	0.126	0.096	0.096	0.095	0.089	0.068	0.063	0.073	-0.002	0.017	0.015	0.032		
P ₇₆																						
LT	0.034	0.051	0.038	0.047	0.010	-0.003	-0.005	-0.014	-0.027	-0.037	-0.036	-0.038	-0.019	-0.026	-0.027	-0.024	-0.005	-0.001	-0.004	0.002		
MT	0.210	0.191	0.195	0.150	0.019	0.024	0.031	0.018	0.118	0.093	0.096	0.093	0.098	0.078	0.075	0.082	0.010	0.027	0.027	0.041		
UT	0.239	0.198	0.204	0.187	-0.174	-0.136	-0.134	-0.150	0.037	0.016	0.018	-0.001	0.018	0.005	0.001	-0.013	-0.025	-0.018	-0.015	-0.021		
P ₅₁																						
LT	0.057	0.069	0.059	0.070	0.009	-0.007	-0.005	-0.012	-0.028	-0.037	-0.035	-0.035	-0.006	-0.008	-0.010	-0.007	0.002	0.005	0.003	0.007		
MT	0.265	0.177	0.159	0.118	0.074	0.062	0.066	0.088	0.145	0.087	0.080	0.099	0.151	0.102	0.093	0.115	0.053	0.056	0.059	0.075		
UT	0.193	0.131	0.103	0.081	-0.019	-0.009	-0.013	0.005	0.059	0.018	0.013	0.022	0.056	0.018	0.014	0.020	0.010	0.005	0.009	0.012		
LS	-0.117	-0.162	-0.140	-0.214	-0.053	-0.075	-0.063	-0.052	-0.051	-0.072	-0.060	-0.054	-0.009	-0.027	-0.010	-0.008	0.060	0.041	0.039	0.056		

Note: The effects of spatial subsampling are calculated as the difference between decadal trends for a subset of radiosounding stations artificially selected (from 20 to 100 stations) versus the decadal trends for the complete dataset, for each of different regression methods in the Northern Hemisphere (NH, $20^{\circ}N < \text{latitude} < 90^{\circ}N$), only due to limited data availability in the Tropics and SH. Decadal trend differences statistically significant with 95% confidence intervals, that is, differences that passed the WMW test at 0.05 level, are reported in bold.

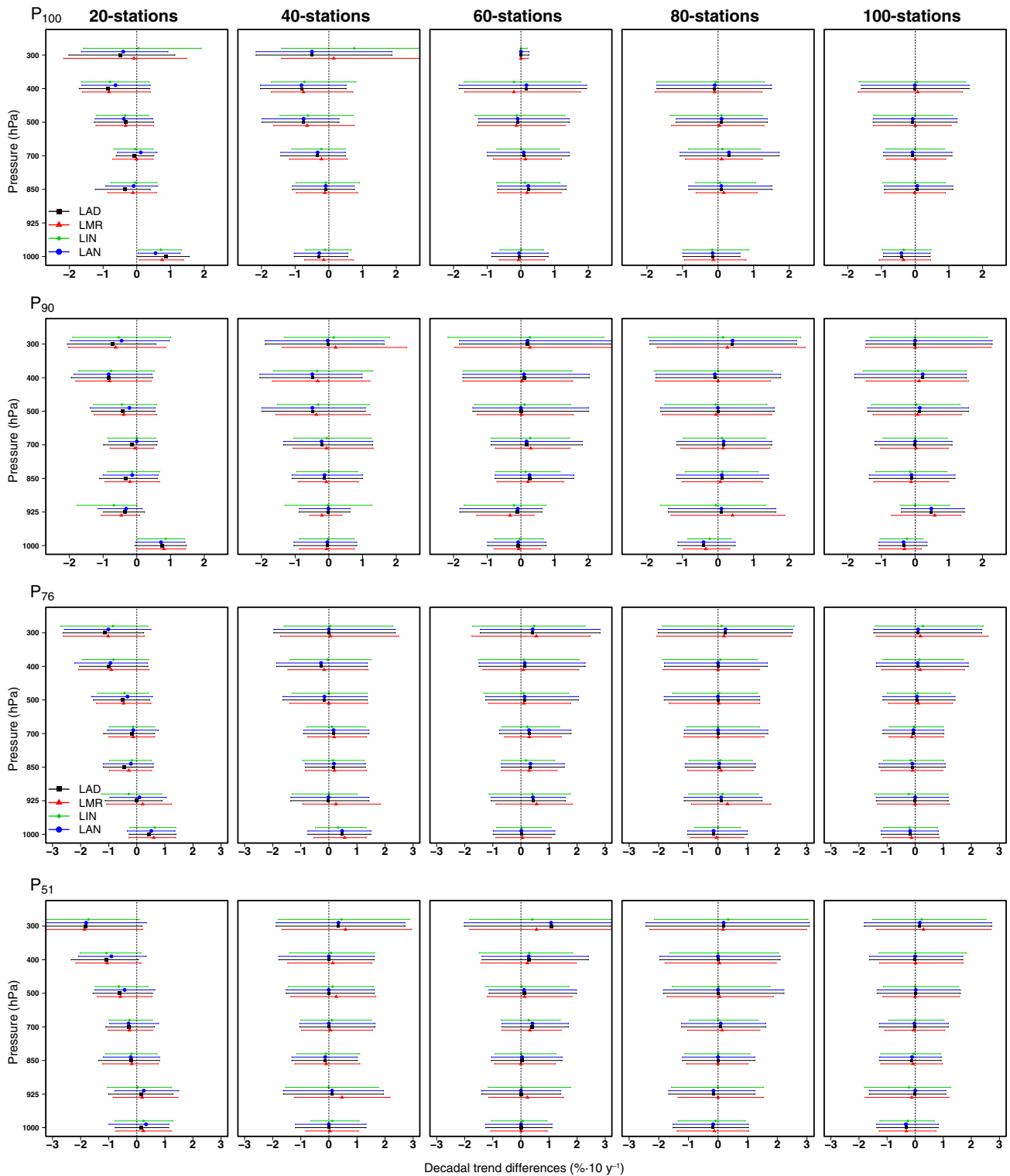


FIGURE 7 Same as Figure 6 but for relative humidity decadal trends (%·decade⁻¹) [Colour figure can be viewed at wileyonlinelibrary.com]

estimation of trends using radiosounding data, the separate roles of temporal and spatial sampling effects due to the missing data must be considered. The debate related to the approach to consider for reducing biases due to sampling errors has been ongoing for several years, and it

is a topic of several papers in literature (e.g., Santer *et al.*, 2000; Rosen *et al.*, 2003; Agudelo and Curry, 2004; Free and Seidel, 2005; McCarthy, 2008). Most of these papers address spatial sampling errors with regard to several factors (i.e., global datasets considered, approach

TABLE 5 Same as Table 4 but for relative humidity (%decade⁻¹) in the middle troposphere (MT) and in the lower troposphere (LT)

N°	20 stations			40 stations			60 stations			80 stations			100 stations							
	Atmospheric layers			LAN	LMR	LAD	LAN	LMR	LAD	LAN	LMR	LAD	LAN	LMR	LAD	LAN	LMR	LAD		
P ₁₀₀																				
LT	-0.019	-0.077	-0.022	-0.055	-0.194	-0.176	-0.190	-0.330	0.132	0.183	0.160	0.247	0.102	0.126	0.106	0.169	-0.002	0.000	-0.009	0.011
MT	-0.282	-0.464	-0.356	-0.495	0.193	0.114	0.062	-0.429	-0.318	-0.183	-0.195	-0.133	-0.249	-0.167	-0.178	-0.144	0.002	0.066	0.051	0.030
P ₉₀																				
LT	-0.050	-0.088	-0.042	-0.043	-0.083	-0.065	-0.081	-0.187	0.132	0.185	0.161	0.241	-0.031	-0.005	-0.021	-0.008	-0.099	-0.089	-0.096	-0.099
MT	-0.318	-0.398	-0.381	-0.399	0.238	0.251	0.210	-0.065	0.024	0.167	0.138	0.277	-0.145	-0.103	-0.113	-0.120	0.096	0.133	0.119	0.083
P ₇₆																				
LT	-0.147	-0.191	-0.105	-0.129	0.134	0.232	0.250	0.121	0.228	0.303	0.301	0.314	0.039	0.087	0.084	0.045	0.068	-0.031	-0.053	-0.050
MT	-0.720	-0.733	-0.718	-0.721	0.629	0.695	0.652	0.437	0.382	0.495	0.467	0.586	0.226	0.242	0.237	0.217	0.204	0.197	0.189	0.160
P ₅₁																				
LT	-0.140	-0.148	-0.090	-0.063	0.050	0.148	0.156	0.078	0.154	0.205	0.196	0.215	-0.035	-0.024	-0.009	-0.073	-0.112	-0.096	-0.100	-0.113
MT	-1.260	-1.233	-1.231	-1.055	0.655	0.585	0.607	0.417	0.622	0.656	0.656	0.758	0.386	0.339	0.336	0.249	0.289	0.229	0.226	0.149

adopted, network configurations and sizes, etc.). In our study, the existing approaches were extended considering: (a) different sub-networks of stations to quantify the impact on decadal trends and (b) uncertainties due to the missing data and to the choice of regression methods, identifying the minimum amount of IGRA stations, which any user could use without a significant increase of the uncertainty due to the spatial sampling effects.

Our results found that enlarging the number of stations, including those with an acceptable level of missing data, can allow us to reduce the trend estimation uncertainties among different regression methods. In analogy, uncertainties are smaller in regions where data are denser (e.g., NH) and time series with a large amount of missing data are filtered out. This can be because missing data within the time series showing no significant trends increases the probability of committing a type-2 error (Schlegel and Smit, 2016). Our results also suggest that the use of the P_{76} dataset is related to a small uncertainty in the estimation of trends due to the choice of the regression method, while the P_{51} dataset may increase the uncertainty up to $0.4 \text{ K}\cdot\text{da}^{-1}$. Uncertainties also critically depend on the coupling of the number of stations selected for an application and the amount of missing data. Selecting a sufficient number of stations (e.g., at least 60) with a limited amount of missing data (e.g., at least P_{76}) can reduce the noise coming from an inappropriate spatial and temporal sampling and, therefore, the uncertainties due to the choice of the trend estimation method. For example, for RH, 40 stations with the coverage of the P_{76} dataset should be selected at minimum.

There are some limitations to this work. First, the dataset selected in this paper has been investigated without considering restrictions on the persistence and the position of the missing data within each time series (i.e., at the beginning, in the end, or in the middle), with the aim to quantify the uncertainty affecting the results of any study based on the IGRA radiosounding data as a function of its data completeness in time. Second, the different subsets of the stations are randomly selected for each P_x and each latitude with the aim to reduce the effects due to the irregular spatial distribution of the stations and their different time coverage at different pressure levels. Nevertheless, this work contributes to increasing the attention and encourages further studies to improve the quantification of uncertainties in climate trends.

ACKNOWLEDGEMENTS

This work was carried out on behalf of the European Union's Copernicus Climate Change Service

implemented by ECMWF. Use of data as stated in the Copernicus licence agreement is acknowledged. Authors would like to gratefully acknowledge the two anonymous reviewers for their helpful comments.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTIONS

Souleymane SY and Fabio Madonna conceived the study, designed the analyses and interpreted the results. Souleymane SY carried out the analyses and wrote the manuscript. Fabio Madonna contributed significantly to the data analysis and the discussions. Marco Rosoldi, Simone Gagliardi, Monica Proto and Gelsomina Pappalardo contributed ideas and provided valuable feedbacks. Souleymane SY and Emanuele Tramutola collected the data and performed all the computational works to make IGRA data useful.

DATA AVAILABILITY STATEMENT

IGRA datasets are provided to the users under the WMO res.40 by NOAA-NCEI (P.I. Dr. Imke Durré) with the following doi:10.7289/V5X63K0Q.

ORCID

Souleymane SY  <https://orcid.org/0000-0001-8834-7620>

Fabio Madonna  <https://orcid.org/0000-0001-7628-8870>

REFERENCES

- Agudelo, P.A. and Curry, J.A. (2004) Analysis of spatial distribution in tropospheric temperature trends. *Geophysical Research Letters*, 31(L2), 1–5. <https://doi.org/10.1029/2004GL020818>.
- Angell, J.K. and Korshover, J. (1975) Estimate of the global change in tropospheric temperature between 1958 and 1973. *Monthly Weather Review*, 103, 1007–1012. [https://doi.org/10.1175/1520-0493\(1975\)103<1007:EOTGCI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1975)103<1007:EOTGCI>2.0.CO;2).
- Barrodale, I. (1968) L1 approximation and the analysis of data. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 17, 51–57. <https://doi.org/10.2307/2985267>.
- Barrodale, I. and Roberts, F.D.K. (1974) Solution of an overdetermined system of equations in the l_1 norm [F4]. *Communications of the ACM*, 17, 319–320.
- Calitz, M.F. and Rüther, H. (1996) Least absolute deviation (LAD) image matching. *ISPRS Journal of Photogrammetry and Remote Sensing*, 51, 223–229. [https://doi.org/10.1016/0924-2716\(95\)00010-0](https://doi.org/10.1016/0924-2716(95)00010-0).
- Choi, S.W. (2009). The effect of outliers on regression analysis: regime type and foreign direct investment. *Quarterly Journal of Political Science*, 4(2), 153–165. <https://doi.org/10.1561/100.00008021>.
- Christy, J.R. and Goodridge, J.D. (1995) Precision global temperatures from satellites and urban warming effects of non-satellite data. *Atmospheric Environment, International Conference on Sustainable Development Strategies and Global/Regional/Local Impacts on Atmospheric Composition*

- and *Climate*, 29, 1957–1961 [https://doi.org/10.1016/1352-2310\(94\)00240-L](https://doi.org/10.1016/1352-2310(94)00240-L).
- Christy, J.R., Spencer, R.W., Norris, W.B., Braswell, W.D. and Parker, D.E. (2003) Error estimates of version 5.0 of MSU–AMSU bulk atmospheric temperatures. *Journal of Atmospheric Oceanic Technology*, 20, 613–629 [https://doi.org/10.1175/1520-0426\(2003\)20<613:EEOVOM>2.0.CO;2](https://doi.org/10.1175/1520-0426(2003)20<613:EEOVOM>2.0.CO;2).
- Dai, A., Wang, J., Thorne, P.W., Parker, D.E., Haimberger, L. and Wang, X.L. (2010) A new approach to homogenize daily radiosonde humidity data. *Journal of Climate*, 24, 965–991. <https://doi.org/10.1175/2010JCLI3816.1>.
- Dervilis, N., Worden, K. and Cross, E.J. (2015) On robust regression analysis as a means of exploring environmental and operational conditions for SHM data. *Journal of Sound and Vibration*, 347, 279–296. <https://doi.org/10.1016/j.jsv.2015.02.039>.
- Dirksen, R.J., Sommer, M., Immler, F.J., Hurst, D.F., Kivi, R. and Vömel, H. (2014) Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde. *Atmospheric Measurement Techniques*, 7, 4463–4490. <https://doi.org/10.5194/amt-7-4463-2014>.
- Durre, I., Vose, R.S. and Wuertz, D.B. (2006) Overview of the integrated global radiosonde archive. *Journal of Climate*, 19, 53–68. <https://doi.org/10.1175/JCLI3594.1>.
- Durre, I., Yin, X., Vose, R.S., Applequist, S. and Arnfield, J. (2018) Enhancing the data coverage in the integrated global radiosonde archive. *Journal of Atmospheric Oceanic Technology*, 35, 1753–1770. <https://doi.org/10.1175/JTECH-D-17-0223.1>.
- Ferreira, A.P., Nieto, R. and Gimeno, L. (2019) Completeness of radiosonde humidity observations based on the integrated global radiosonde archive. *Earth System Science Data*, 11, 603–627. <https://doi.org/10.5194/essd-11-603-2019>.
- Finger, F.G., Nagatani, R.M., Gelman, M.E., Long, C.S. and Miller, A.J. (1995) Consistency between variations of ozone and temperature in the stratosphere. *Geophysical Research Letters*, 22, 3477–3480. <https://doi.org/10.1029/95GL02786>.
- Finger, R. (2010) Revisiting the evaluation of robust regression techniques for crop yield data detrending. *American Journal of Agricultural Economics*, 92, 205–211. <https://doi.org/10.1093/ajae/aap021>.
- Free, M. and Seidel, D.J. (2005) Causes of differing temperature trends in radiosonde upper air data sets. *Journal of Geophysical Research Atmospheres*, 110(D07101), 1–14. <https://doi.org/10.1029/2004JD005481>.
- Gaffen, D.J. (1994) Temporal inhomogeneities in radiosonde temperature records. *J. Geophys. Res. Atmospheres*, 99, 3667–3676. <https://doi.org/10.1029/93JD03179>.
- Gaffen, D.J. (1998) Falling satellites, rising temperatures? *Nature*, 394, 615–616. <https://doi.org/10.1038/29163>.
- Gaffen, D.J., Sargent, M.A., Habermann, R.E. and Lanzante, J.R. (2000) Sensitivity of tropospheric and stratospheric temperature trends to radiosonde data quality. *Journal of Climate*, 13, 1776–1796 [https://doi.org/10.1175/1520-0442\(2000\)013<1776:SOTAST>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1776:SOTAST>2.0.CO;2).
- Hartmann, D.L., Tank, A.M.K., Rusticucci, M., Alexander, L.V., Brönnimann, S., Charabi, Y.A.R., Dentener, F.J., Dlugokencky, E.J., Easterling, D.R. and Kaplan, A. (2013) Observations: atmosphere and surface. In: *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press, pp. 159–254.
- Helsel, D.R. and Hirsch, R.M. (1992) *Statistical Methods in Water Resources*. Netherlands: Elsevier.
- Hoaglin, D.C., Iglewicz, B. and Tukey, J.W. (1986) Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81, 991–999. <https://doi.org/10.1080/01621459.1986.10478363>.
- Hoaglin, D.C., Mosteller, F. and Tukey, J.W. (1983) *Understanding Robust and Exploratory Data Analysis*. New York: John Wiley & Sons.
- Ingleby, B. (2017). An assessment of different radiosonde types 2015/2016. Technical Memorandum 71.
- Karl, T.R., Hassol, S.J., Miller, C.D., and Murray, W.L. (2006). Temperature trends in the lower atmosphere: steps for understanding and reconciling differences. *A report by the climate change science program and subcommittee on global change research*. Washington, DC: Climate Change Science Program, p. 180.
- Koller, M. and Stahel, W.A. (2011) Sharpening Wald-type inference in robust regression for small samples. *Computational Statistics and Data Analysis*, 55, 2504–2515. <https://doi.org/10.1016/j.csda.2011.02.014>.
- Ladstädter, F., Steiner, A.K., Foelsche, U., Haimberger, L., Tavolato, C. and Kirchengast, G. (2011) An assessment of differences in lower stratospheric temperature records from (a) MSU, radiosondes, and GPS radio occultation. *Atmospheric Measurement Techniques*, 4, 1965–1977. <https://doi.org/10.5194/amt-4-1965-2011>.
- Lanzante, J.R. (1996) Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical Radiosonde Station data. *International Journal of Climatology*, 16, 1197–1226 [https://doi.org/10.1002/\(SICI\)1097-0088\(199611\)16:11<1197::AID-JOC89>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-0088(199611)16:11<1197::AID-JOC89>3.0.CO;2-L).
- Lanzante, J.R., Klein, S.A. and Seidel, D.J. (2003) Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology. *Journal of Climate*, 16, 224–240 [https://doi.org/10.1175/1520-0442\(2003\)016<0224:THOMRT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<0224:THOMRT>2.0.CO;2).
- Lettenmaier, D.P., Wood, E.F. and Wallis, J.R. (1994) Hydro-climatic trends in the continental united states, 1948–88. *Journal of Climate*, 7(4), 586–607. [https://doi.org/10.1175/1520-0442\(1994\)007<0586:HCTITC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<0586:HCTITC>2.0.CO;2).
- Madonna, F., Kivi, R., Dupont, J.-C., Ingleby, B., Fujiwara, M., Romanens, G., Hernandez, M., Calbet, X., Rosoldi, M., Giunta, A., Karppinen, T., Iwabuchi, M., Hoshino, S., von Rohden, C. and Thorne, P.W. (2020a) Use of automatic radiosonde launchers to measure temperature and humidity profiles from the GRUAN perspective. *Atmospheric Measurement Techniques*, 13, 3621–3649. <https://doi.org/10.5194/amt-13-3621-2020>.
- Madonna, F., Tramutola, E., Sy, S., Serva, F., Proto, M., Rosoldi, M., Gagliardi, S., Amato, F., Marra, F., Fassò, A., Gardiner, T. and Thorne, P.W. (2020b) Radiosounding HARMonization (RHARM): a new homogenized dataset of radiosounding temperature, humidity and wind profiles with uncertainty. *Earth System Science Data Discussions*, 1–38. <https://doi.org/10.5194/essd-2020-183>, in review.
- McCarthy, M.P. (2008) Spatial sampling requirements for monitoring upper-air climate change with radiosondes. *International Journal of Climatology*, 28, 985–993. <https://doi.org/10.1002/joc.1611>.
- Miloshevich, L.M., Vömel, H., Whiteman, D.N., Lesht, B.M., Schmidlin, F.J. and Russo, F. (2006) Absolute accuracy of water vapor measurements from six operational radiosonde types launched during AWEX-G and implications for AIRS validation. *Journal of Geophysical Research Atmospheres*, 111(D09), 1–25. <https://doi.org/10.1029/2005JD006083>.

- Mudelsee, M. (2019) Trend analysis of climate time series: a review of methods. *Earth Science Reviews*, 190, 310–322. <https://doi.org/10.1016/j.earscirev.2018.12.005>.
- Oakley, T. (1993). *Report by the Rapporteur on Radiosonde Compatibility Monitoring. Part B - Compatibility of Radiosonde Geopotential Measurements 1990, 1991, 1992. IOM Report No. 56, WMO/TD No. 587.* Geneva. Available at: <https://library.wmo.int> [accessed 25th February 2020].
- Oort, A.H. (1978) Adequacy of the Rawinsonde network for global circulation studies tested through numerical model output. *Monthly Weather Review*, 106, 174–195. [https://doi.org/10.1175/1520-0493\(1978\)106](https://doi.org/10.1175/1520-0493(1978)106).
- Parker, D.E. (1985) On the detection of temperature changes induced by increasing atmospheric carbon dioxide. *Quarterly Journal of the Royal Meteorological Society*, 111, 587–601. <https://doi.org/10.1002/qj.49711146816>.
- Rice, J. and White, J. (1964) Norms for smoothing and estimation. *SIAM Review*, 6, 243–256. <https://doi.org/10.1137/1006061>.
- Rosen, R.D., Henderson, J.M. and Salstein, D.A. (2003) Sensitivity of continental-scale climate trend estimates to the distribution of radiosondes over North America. *Journal of Atmospheric Oceanic Technology*, 20, 262–268 [https://doi.org/10.1175/1520-0426\(2003\)020<0262:SOCST>2.0.CO;2](https://doi.org/10.1175/1520-0426(2003)020<0262:SOCST>2.0.CO;2).
- Rousseeuw, P.J. and Hubert, M. (2011) Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1, 73–79. <https://doi.org/10.1002/widm.2>.
- Santer, B.D., Wigley, T.M.L., Boyle, J.S., Gaffen, D.J., Hnilo, J.J., Nychka, D., Parker, D.E. and Taylor, K.E. (2000) Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *Journal of Geophysical Research Atmospheres*, 105, 7337–7356. <https://doi.org/10.1029/1999JD901105>.
- Schlegel, R.W. and Smit, A.J. (2016) Climate change in coastal waters: time series properties affecting trend estimation. *Journal of Climate*, 29, 9113–9124. <https://doi.org/10.1175/JCLI-D-16-0014.1>.
- Siegel, A.F. and Benson, R.H. (1982) A robust comparison of biological shapes. *Biometrics*, 38, 341–350. <https://doi.org/10.2307/2530448>.
- Simmons, A.J., Poli, P., Dee, D.P., Berrisford, P., Hersbach, H., Kobayashi, S. and Peubey, C. (2014) Estimating low-frequency variability and trends in atmospheric temperature using ERA-interim: atmospheric temperature variability and trends using ERA-interim. *Quarterly Journal of the Royal Meteorological Society*, 140, 329–353. <https://doi.org/10.1002/qj.2317>.
- Susanti, Y., Pratiwi, H., H, S.S and Liana, T. (2014) M estimation, s estimation, and mm estimation in robust regression. *International Journal of Pure and Applied Mathematics*, 91, 349–360.
- Sy, S. and Quesada, B. (2020) Anthropogenic land cover change impact on climate extremes during the 21st century. *Environmental Research Letters*, 15(3), 1–9. <https://doi.org/10.1088/1748-9326/ab702c>.
- Theil, H. (1950) A rank-invariant method of linear and polynomial regression analysis I,II and III. *Proceedings of the Koninklijke Nederlandse Akademie Wetenschappen, Series A Mathematical Sciences*, 53, 1397–1412.
- Thorne, P.W., Parker, D.E., Tett, S.F.B., Jones, P.D., McCarthy, M., Coleman, H. and Brohan, P. (2005) Revisiting radiosonde upper air temperatures from 1958 to 2002. *Journal of Geophysical Research Atmospheres*, 110(D18), 1–17. <https://doi.org/10.1029/2004JD005753>.
- Thorne, W.P., Madonna, F., Schulz, J., Oakley, T., Ingleby, B., Rosoldi, M., Tramutola, E., Arola, A., Buschmann, M., Mikalsen, C.A., Davy, R., Voces, C., Kreher, K., De Maziere, M. and Pappalardo, G. (2017) Making better sense of the mosaic of environmental measurement networks: a system-of-systems approach and quantitative assessment. *Geoscientific Instrumentation, Methods and Data Systems*, 6, 453–472. <https://doi.org/10.5194/gi-6-453-2017>.
- Trenberth, K.E. and Olson, J.G. (1991) Representativeness of a 63-station network for depicting climate changes. In: Schlesinger, M.E. (Ed.) *Developments in Atmospheric Science, Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations*. Amsterdam: Elsevier, pp. 249–259. <https://doi.org/10.1016/B978-0-444-88351-3.50023-4>.
- Weatherhead, E.C., Reinsel, G.C., Tiao, G.C., Meng, X.-L., Choi, D., Cheang, W.-K., Keller, T., DeLuisi, J., Wuebbles, D.J., Kerr, J.B., Miller, A.J., Oltmans, S.J. and Frederick, J.E. (1998) Factors affecting the detection of trends: statistical considerations and applications to environmental data. *Journal of Geophysical Research Atmospheres*, 103, 17149–17161. <https://doi.org/10.1029/98JD00995>.
- Weatherhead, E.C., Stevermer, A.J. and Schwartz, B.E. (2002) Detecting environmental changes and trends. *Physics and Chemistry of the Earth Parts ABC*, 27, 399–403. [https://doi.org/10.1016/S1474-7065\(02\)00019-0](https://doi.org/10.1016/S1474-7065(02)00019-0).
- Whiteman, D.N., Vermeesch, K.C., Oman, L.D. and Weatherhead, E.C. (2011) The relative importance of random error and observation frequency in detecting trends in upper tropospheric water vapor. *Journal of Geophysical Research Atmospheres*, 116(D9), 1–25. <https://doi.org/10.1029/2011JD016610>.
- Wilks, D.S. (2011) *Statistical Methods in the Atmospheric Sciences*. Amsterdam: Academic Press.
- WMO (Ed.). (1977) *Abridged Report with Resolutions*. Geneva: Secretariat of the World Meteorological Organization.
- WMO. (1996) *Guide to Meteorological Instruments and Methods of Observation*, 6th edition. Geneva: World Meteorological Organization.
- Wong, R.K.W., Schneider, C., Jr. and Mielke, P.W. (1989) Geometric consistency for regression model estimation and testing in climatology and meteorology. *Atmosphere-Ocean*, 27, 508–520 <https://doi.org/10.1080/07055900.1989.9649349>.
- Zaman, A., Rousseeuw, P.J. and Orhan, M. (2001) Econometric applications of high-breakdown robust regression techniques. *Economics Letters*, 71, 1–8. [https://doi.org/10.1016/S0165-1765\(00\)00404-3](https://doi.org/10.1016/S0165-1765(00)00404-3).

How to cite this article: SY S, Madonna F, Rosoldi M, *et al.* Sensitivity of trends to estimation methods and quantification of subsampling effects in global radiosounding temperature and humidity time series. *Int J Climatol.* 2020;1–23. <https://doi.org/10.1002/joc.6827>