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

Souleymane SY, Fabio Madonna, Monica Proto, Marco Rosoldi, Emanuele Tramutola and Gelsomina Pappalardo

Italian National Research Council (CNR), Institute of Methodologies for Environmental Analysis (CNR-IMAA), Zona Industriale 85050 Tito Scalo (Potenza) – Italy

EGU General Assembly 2020

"Sharing Geoscience Online"

Contact informations:
Dr. Souleymane SY
Email: souleymane.sy@imaa.cnr.it

1. Introduction

- Trend estimation is important for climate change detection. Its inaccurate calculation may lead to incorrect conclusions about the current state and future evolution of the climate.
- These sources of uncertainty in the trend estimation must be added to other contributions like the trend sensitivity to the choice of regression methods and those due to measurements subsampling both in time, due to gaps (e.g. missing data) in the data records and, in space.

Objectives

- 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.
- 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 measurements network

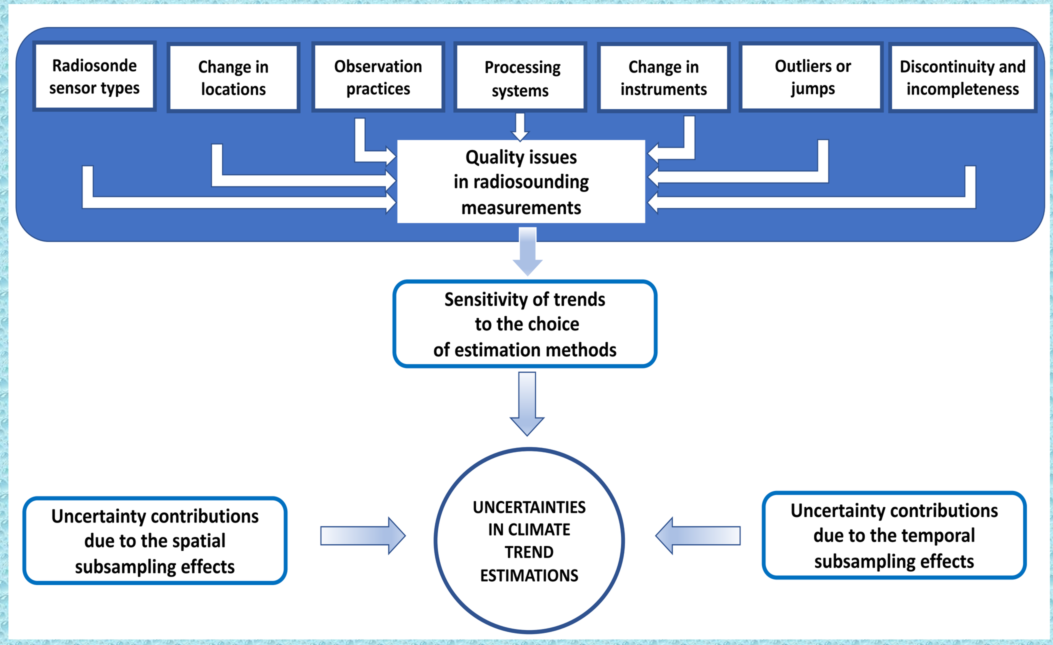

Figure 3: Graphical Abstract

2. Data & Methods

2.1 Radiosonde Data sets:

Radiosonde data set records from the Integrated Global Radiosonde Archive version 2 (hereafter, IGRA) are used in this work (Durré et al., 2018).

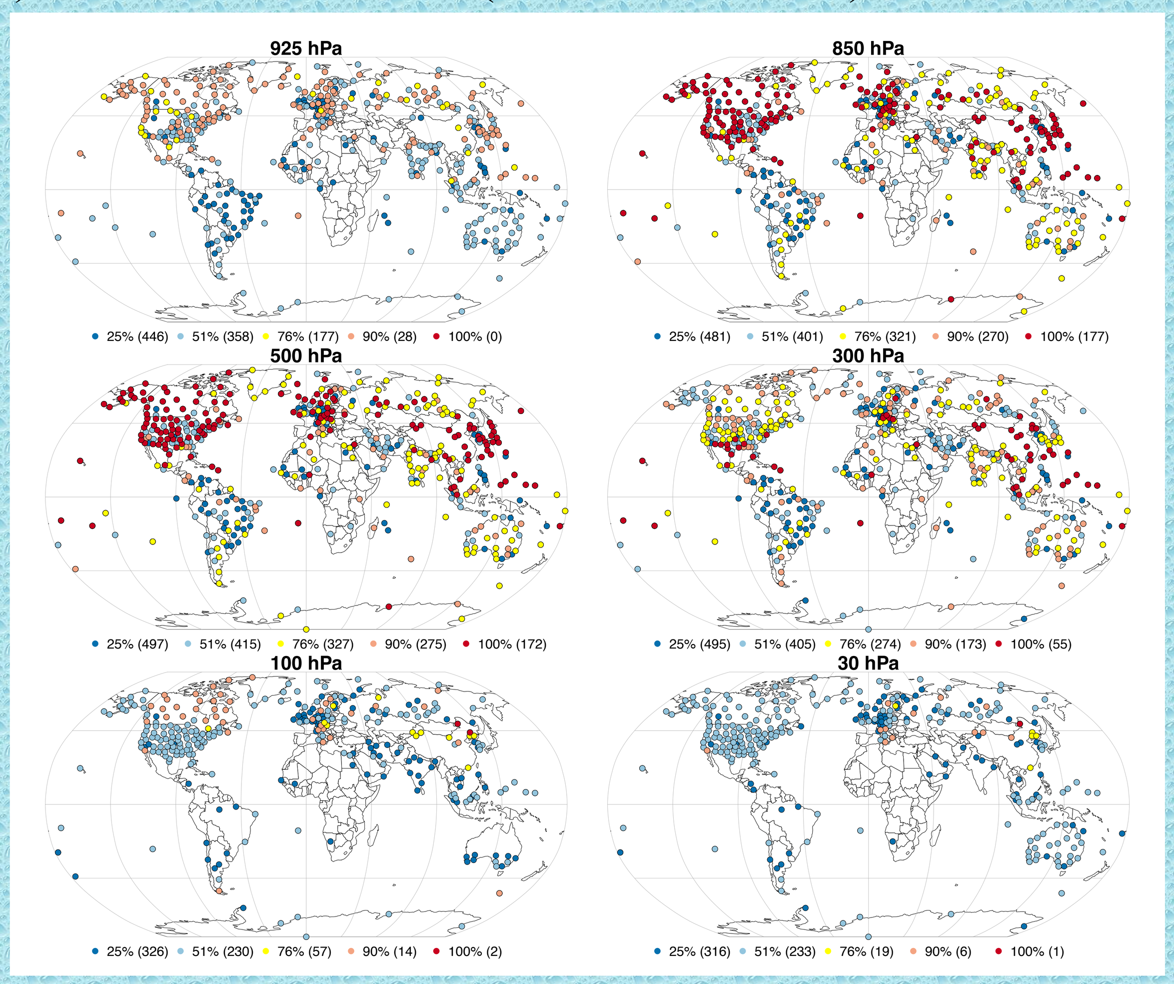

Figure 1: Spatial distribution of radiosounding stations with 0, 5, 10, 20, 30 years of missing data. Missing data are reported as the percentage of months where data are available to the total number of months i.e, X=100%, 90%, 76%, 51% or 25%. Each dot indicates a station, each color indicates the percentage Px 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.

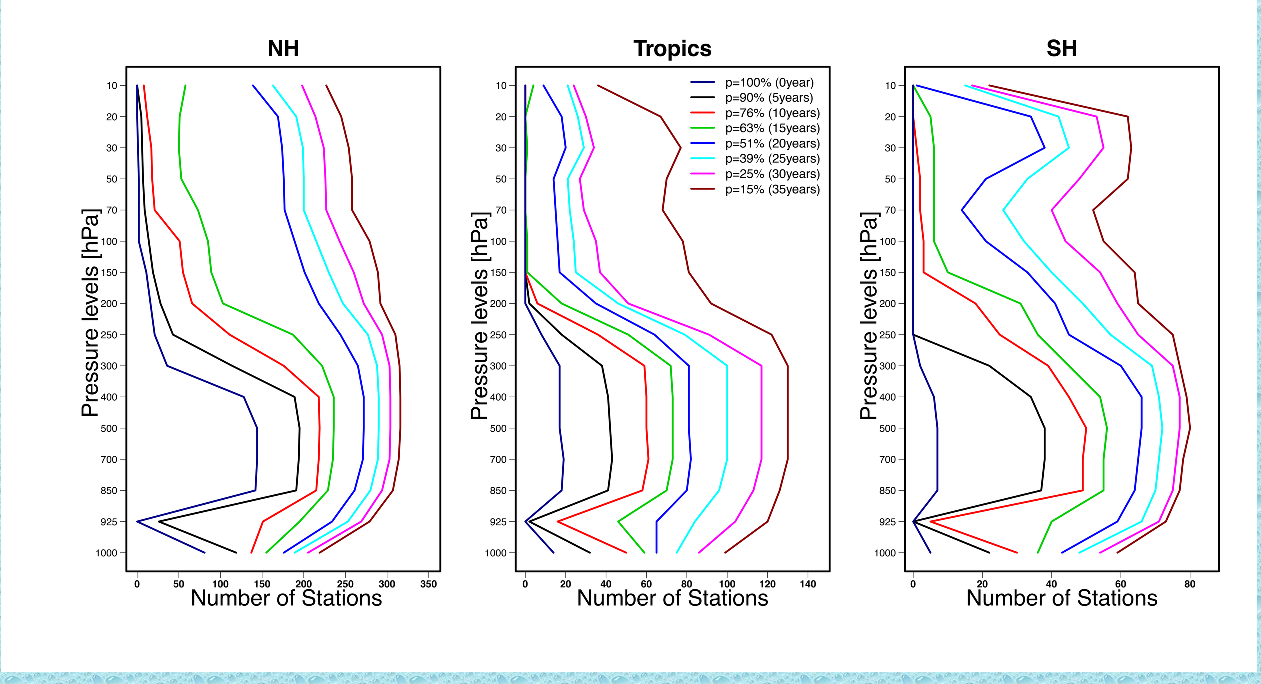

Figure 2: Number of radiosonde stations recording a given percentage P_x of temperature and relative humidity data at each pressure level since 1978 to 2018 for different latitudinal belts. In brackets, the amount of missing data in number of years is reported.

2.2 Sensitivity Analysis

Methods of estimating trends:

- ***Simple linear regression technique (hereafter, LIN),*** a parametric regression method no resistant to outliers based on statistical significance via a T-test.
- ***Lanzante robust linear fitting method (hereafter, LAN)****,* non-parametric regression based on the median of pairwise slopes regression (Lanzante, 1996).
- ***Least Absolute Deviation regression (hereafter, LAD),*** least absolute deviation method based on Barrodale-Roberts (1974) algorithm.
- ***LMROB (hereafter, LMR),*** non-parametric regression method based on MM-estimator for linear regression models (Susanti et al., 2014)

SENSITIVITY OF TREND ESTIMATIONS

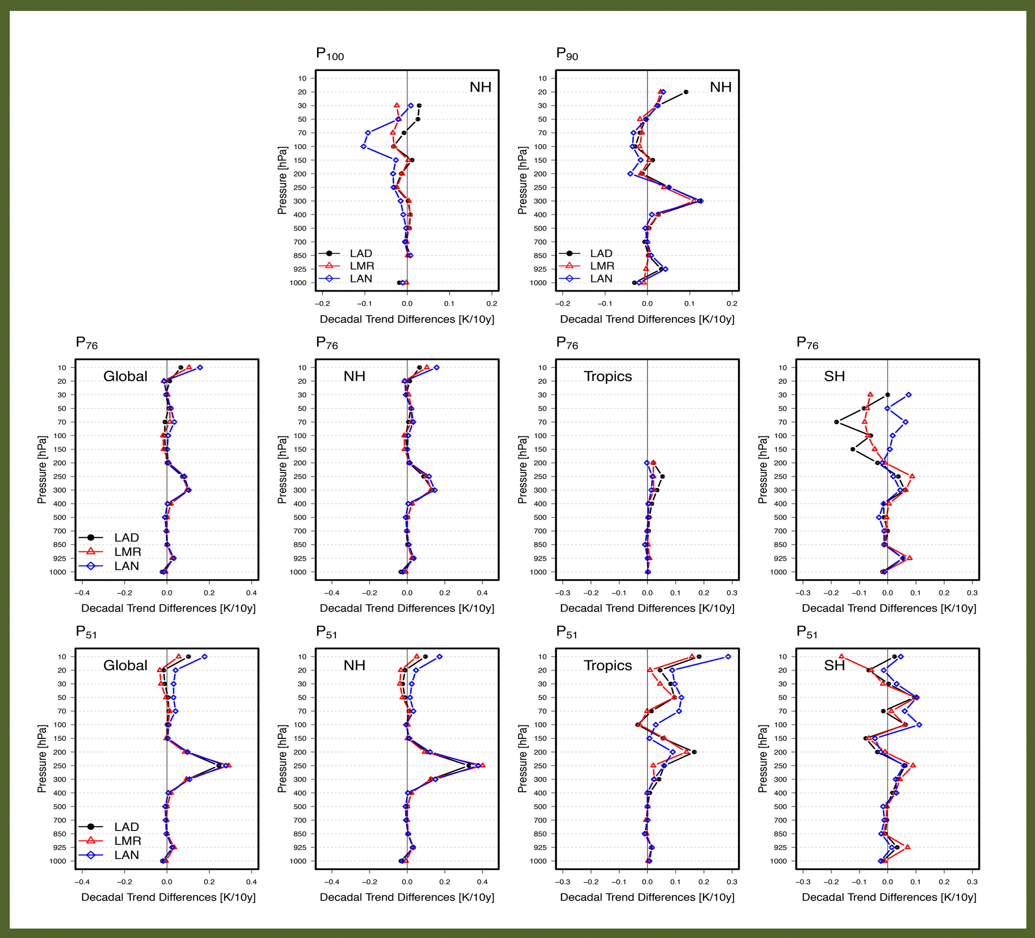

Figure 4: Temperature decadal trend differences (K/decade) estimated between pairs of non-parametric linear regression methods with respect to the simple parametric linear method. Due to the limited observation available, differences calculated based on observations P100 and P90 (top panels) are only shown in the NH.

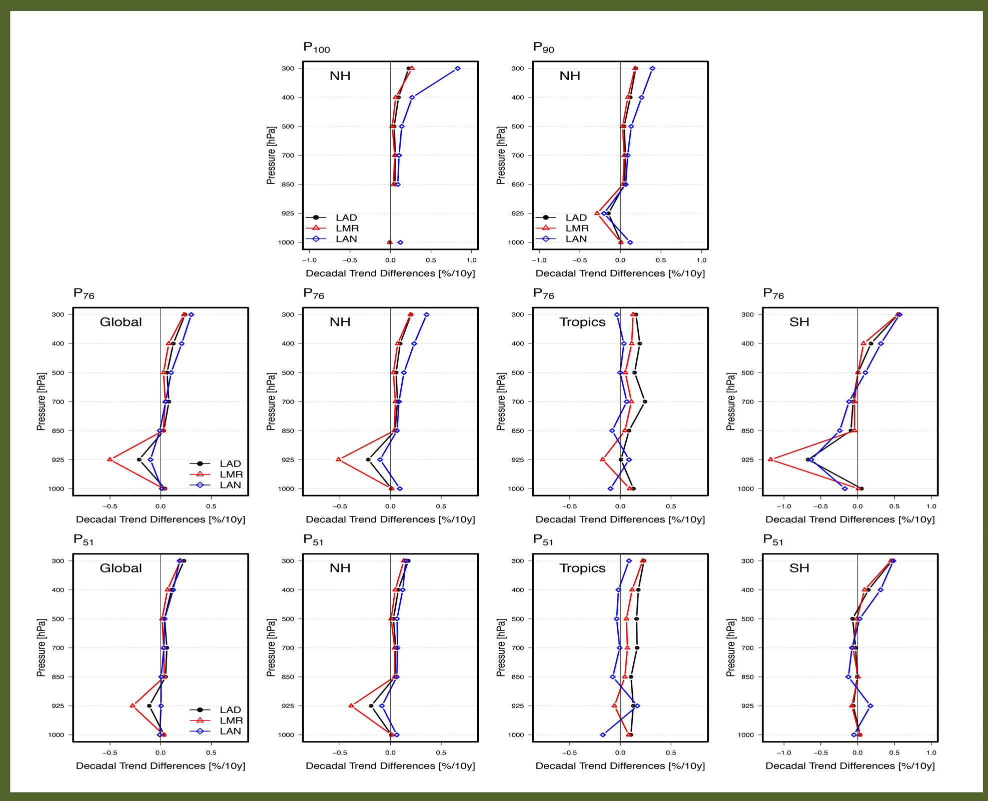

Figure 5. Same as Figure 4 but for RH trend differences (%/decade).

TEMPORAL SAMPLING EFFECTS

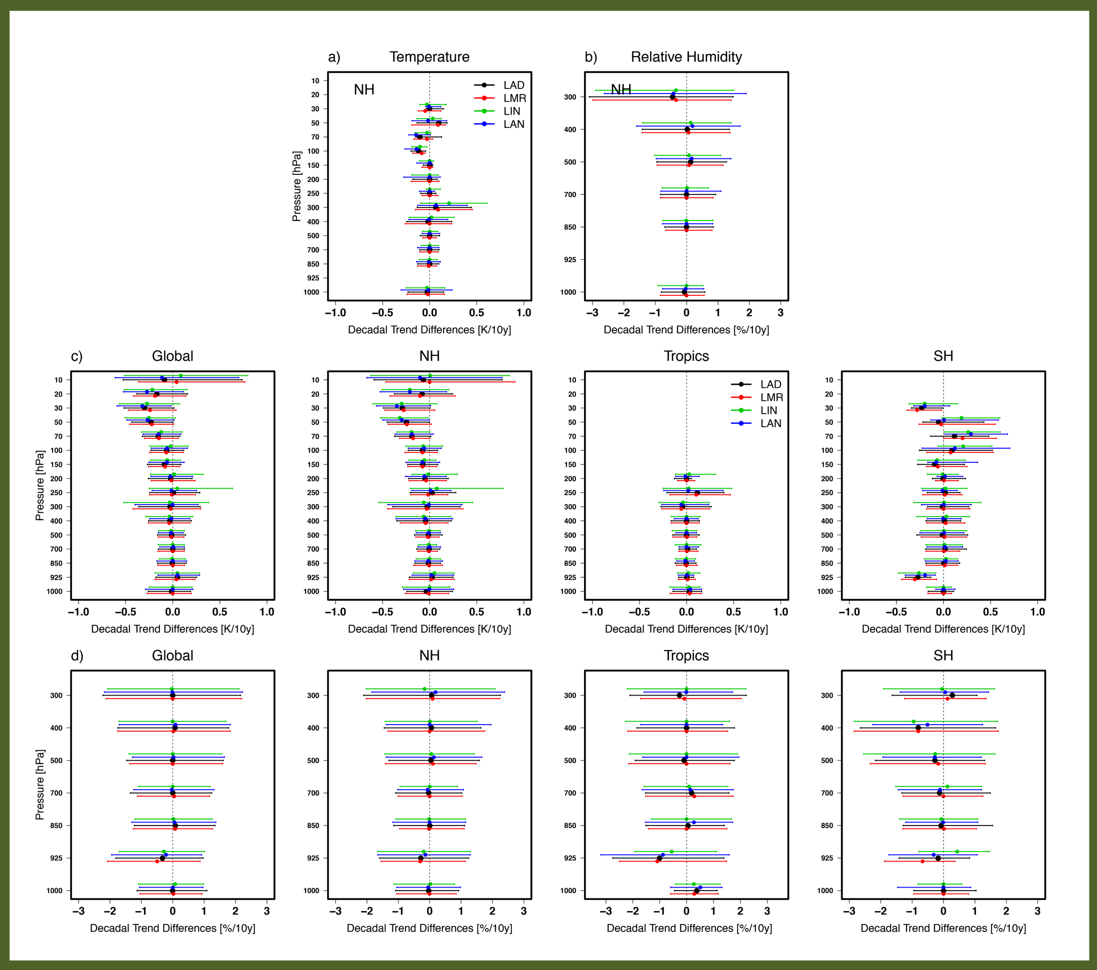

Figure 6. Sensitivity of temperature (K/decade) and relative humidity (%/decade) decadal trends to the missing data effects. The effect is quantified using the differences between decadal trends estimated for two different values of Px (i.e Px1 – Px2). Due to the limited observation available, trend differences calculated between P100 and P90 datasets (i.e. P100−P90) are only shown for the NH. Differences between decadal trends estimated for P76 and for P51 (i.e. P76−P51) are shown in all latitudes. Dots are representative of the median values for each of different regression methods while horizontal bars are representative of the 1st and 3rd quartiles of the corresponding probability distribution

SPATIAL SAMPLING EFFECTS

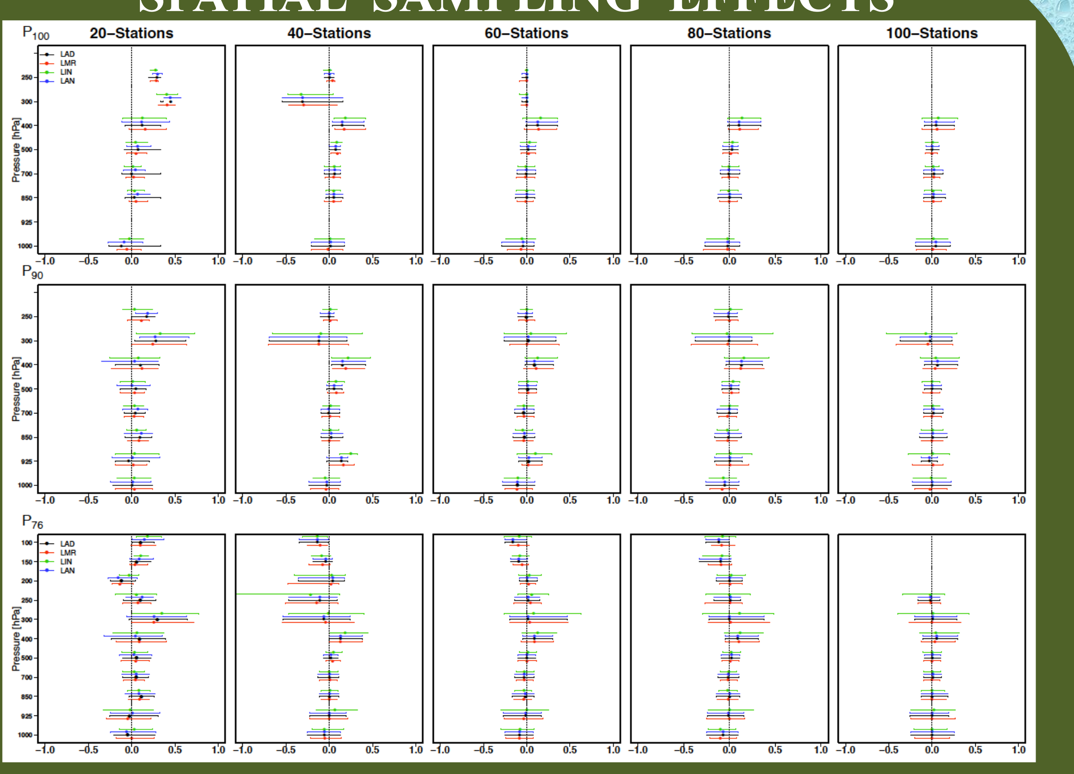

Figure 7: Sensitivity of temperature decadal trends (K/decade) to the spatial subsampling effects. The effects are estimated as the difference between trends estimated from each subset of radiosounding stations artificially selected (ranging from 20 to 100) versus the complete network in the NH only (due to the limited observation availability) at all mandatory pressure levels for P100 (top panels), for P90 (top-middle panels) and for P76 (bottom panels). The dots are representative of the median values for each of different regression methods while horizontal bars are representative of the 1st and 3rd quartiles

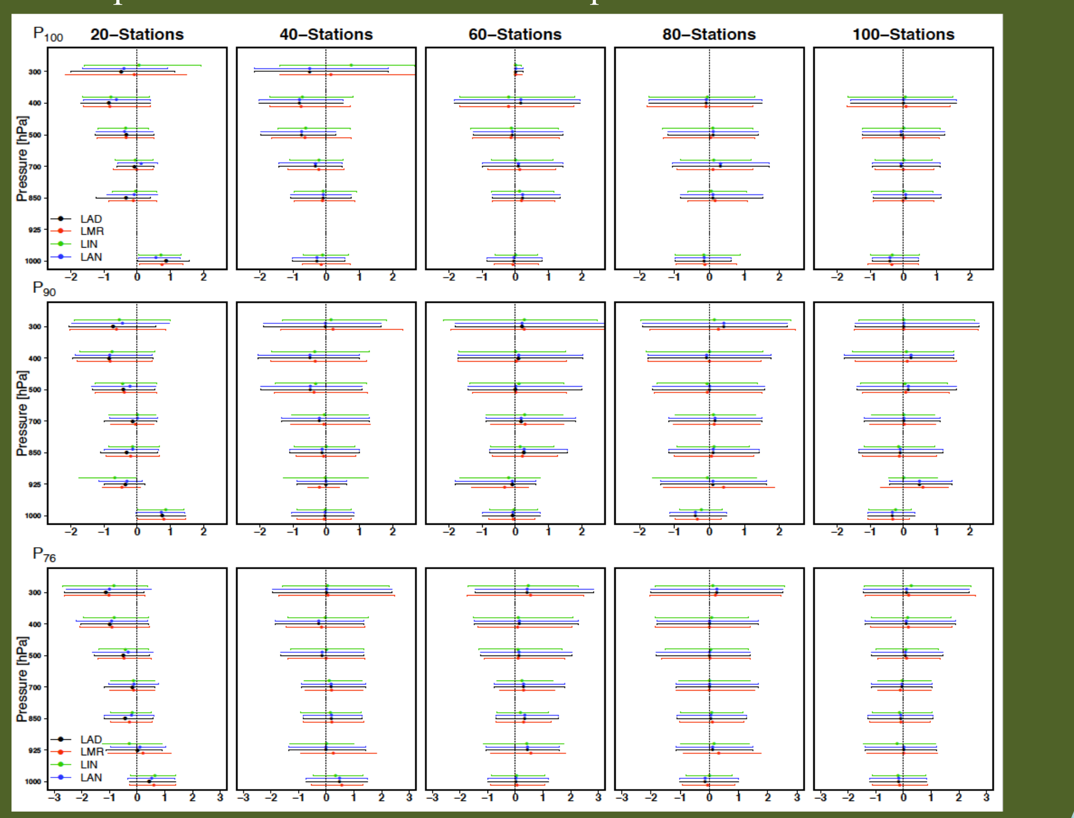

Figure 8: Same as Figure 7 but for relative humidity decadal trends (%/decade).

References:

Durre, I., Yin, X., Vose, R. S., Applequist, S., & Arnfield, J. (2018). Enhancing the Data Coverage in the Integrated Global Radiosonde Archive. *Journal of Atmospheric and Oceanic Technology*, *35*(9), 1753-1770.

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: A Journal of the Royal Meteorological Society, 16(11), 1197-1226.

Acknowledgements:

This work was funded from the Copernicus program under service contract 2017/C3S_311a_Lot3_CNR/SC2. Use of C3S data as stated in the C3S license agreement or Copernicus licence agreement is acknowledged

Discussion & Conclusion:

- Although using complete datasets, differences among regression methods may vary from −0.10K/da to −0.01 K/da below 100 hPa for temperature below 100 hPa and from 0.2%/da to 0.8%/da at 300 hPa for RH. These uncertainties may be significant considering the GCOS requirements (GCOS, 2007) and can be due to several aspects from the non-normality and skewness of the data distribution and the presence of outliers or/and large change-points which may differently affect the different methods used.
- Results show that enlarging the number of stations including those with an acceptable level of missing data can allow 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.
- Concluding, although the completeness of historical radiosounding observations is improving over the years, missing data in the time series are still frequent. This work confirms the need to carry out further comparisons to provide a robust quantification of the uncertainties in the estimation of climate trends.