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

IMPLEMENTED BY ECMWF

Souleymane SY, Fabio Madonna, Monica Proto, Marco Rosoldi, Emanuele Tramutola and Gelsomina Pappalardo
Consiglio Nazionale delle Ricerche - Istituto di Metodologie per l'Analisi Ambientale (CNR-IMAA), Tito Scalo (Potenza), Italy
Board number P49

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

1. Introduction

- Radiosonde profiling in-situ observations have proven to be essential for the study of weather and climate and have been frequently used for the estimation of climate trends.
- 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.
- It is still challenging to provide a robust trend estimations for temperature and humidity from radiosonde data sets because radiosounding time series are affected by several inhomogeneities (due to the changes in the utilized sensor type at different locations)
- 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

- Analyse source uncertainties in the estimation of decadal trends in radiosounding historical time series due to the choice of linear regression methods.
- Provide quantitative estimates of the uncertainty introduced by the spatial and temporal subsampling effects on decadal trends estimations.

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).

RADIOSONDE STATIONS: QUALITY CHECKS

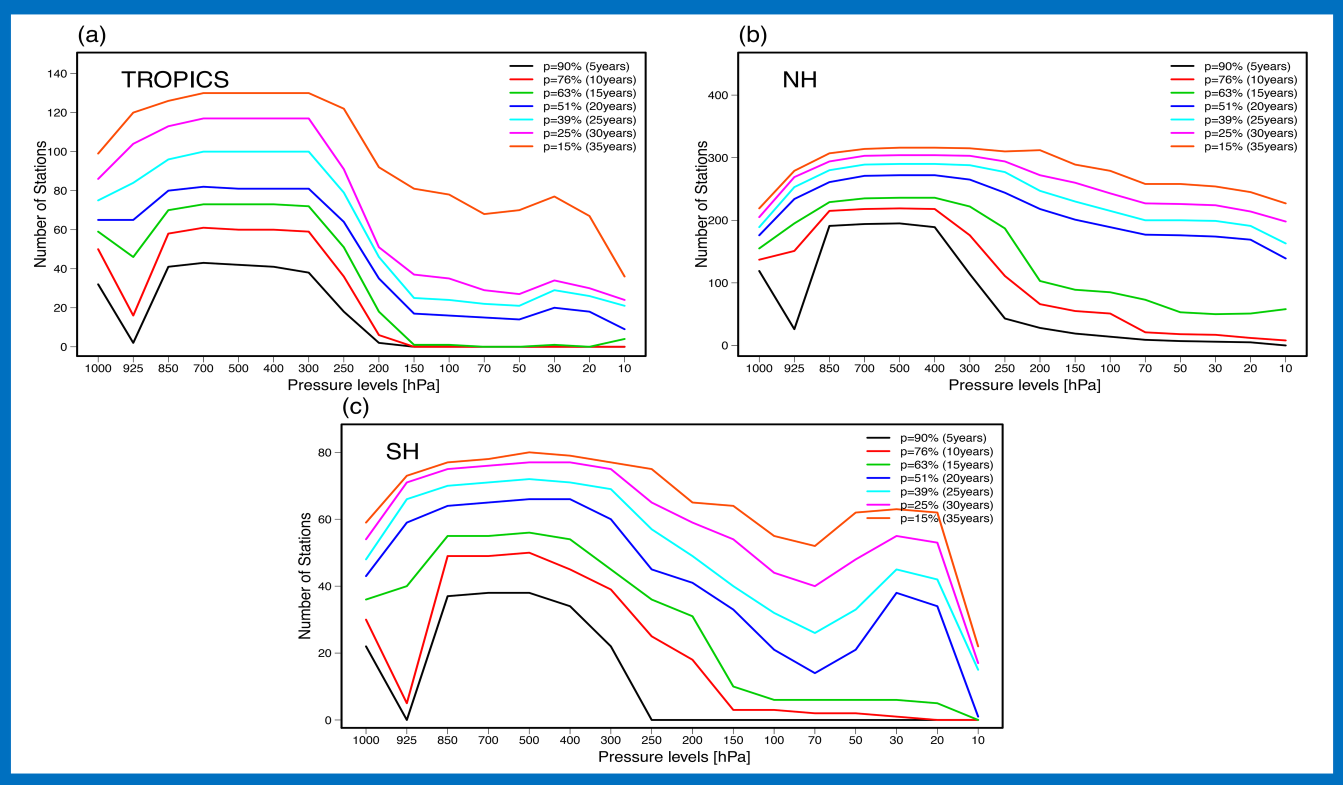

Figure 1. Number of radiosonde stations recording at least a given percentage P_x of temperature and relative humidity monthly data at mandatory pressure levels since 1978 to present time at the Northern hemisphere (NH; latitudes >20° N) (a), the tropics (± 20° of latitudes) (b) and the Southern Hemisphere (SH; latitudes >20° S

2.2 Sensitivity Analysis

Statistical methods

Statistical methods widely used to assess linear regression are based on a number of fundamental assumptions which are often violated in trends estimations.
Violations of the assumptions considered here include the presence of outliers in data set, non-Gaussian behaviour and statistical non-stationarity.

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)

3. Results

SENSITIVITY OF TREND ESTIMATIONS

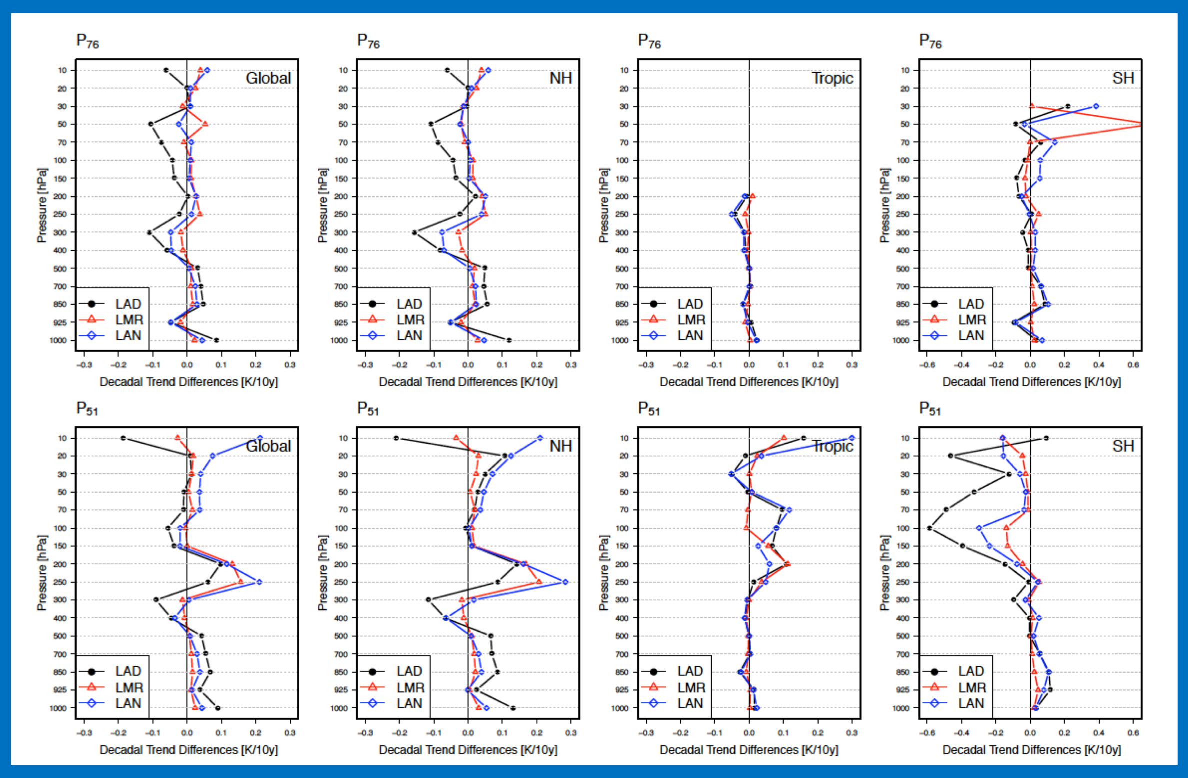

Figure 2: Temperature decadal trend differences (K/decade) estimated between each of the three non-parametric regression methods (LAD, LMR, LAN) and LIN parametric regression method, i.e.: (LAD minus LIN; black curve), (LMR minus LIN; red curve) and (LAN minus LIN; blue curve) for P_{76} (top panels), P_{51} (middle panels) .

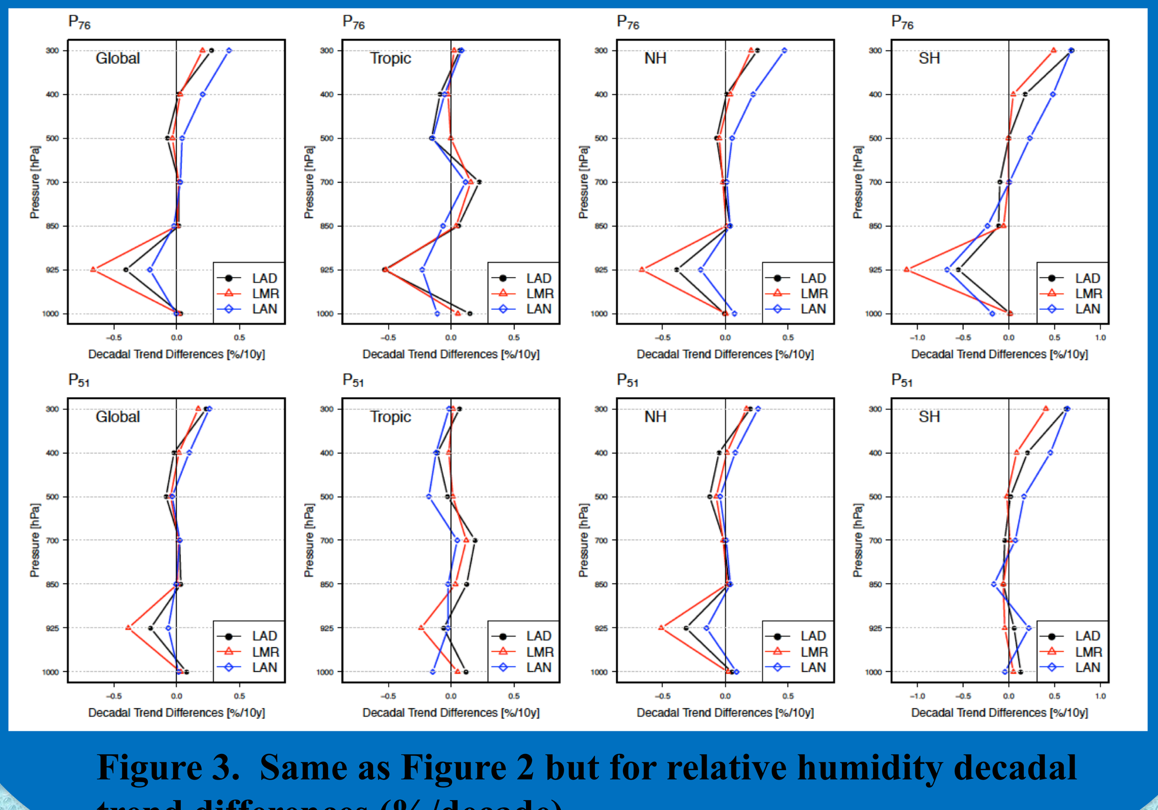

Figure 3. Same as Figure 2 but for relative humidity decadal trend differences (%/decade).

TEMPORAL SAMPLING EFFECTS

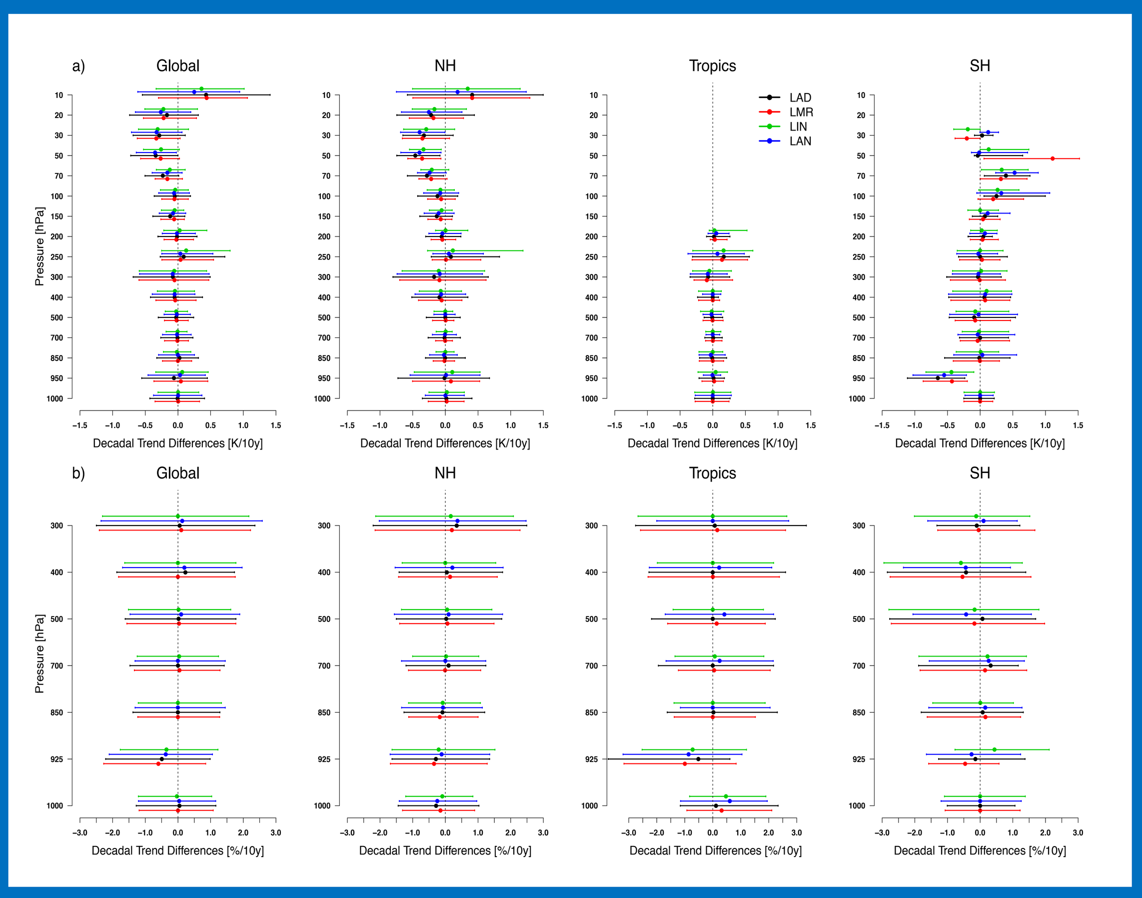

Figure 4. Sensitivity of trends to temporal subsampling effects. Decadal trend differences are estimated for temperature (top panels) and for relative humidity (bottom panels) using radiosonde from P_{76} datasets and P_{51} datasets (i.e., P_{76}-P_{51}) at all latitudinal belts and for each regression method: LAD (black color), LMR (red color), LIN (green color) and LAN (blue color). The dots are representative of the median values for each of different methods while horizontal bars are representative the variability of each method calculated from the 1st and 3rd quartiles of the corresponding probability

SPATIAL SAMPLING EFFECTS

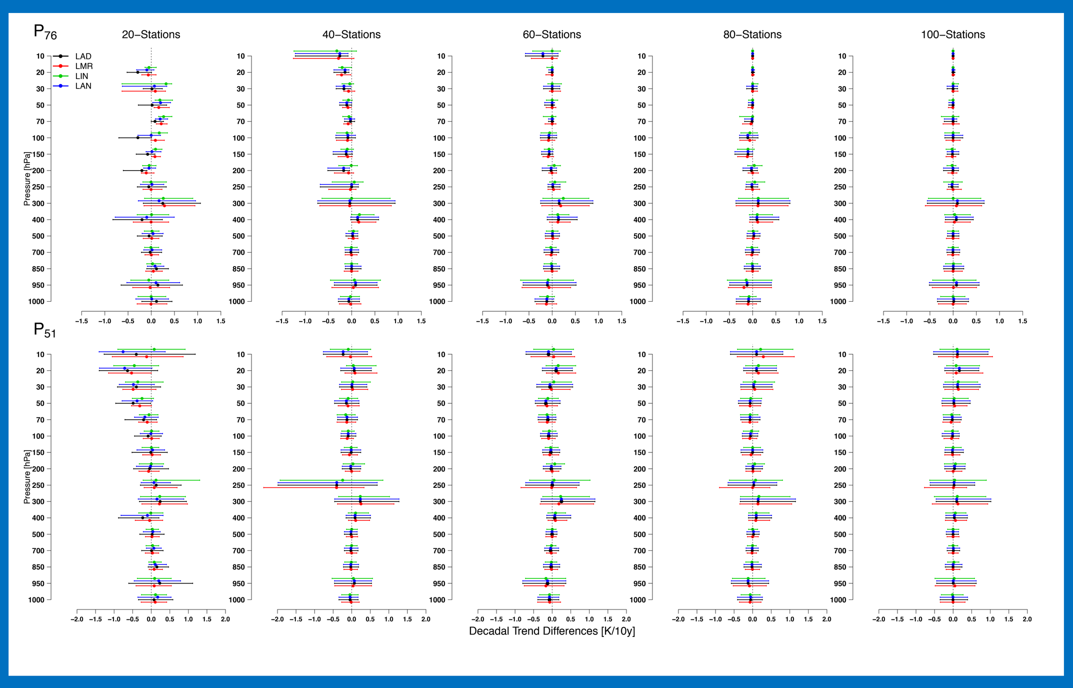

Figure 5: Sensitivity of temperature decadal trends to the spatial subsampling effects. The Spatial subsampling effects are estimated as the difference of trends estimated for different subsets of IGRA radiosonde stations ranging from 20 to 100 correponding to P_{76} (top panels) and P_{51} (bottom panels) with respect to the full dataset of IGRA stations at the NH and at each mandatory pressure level .

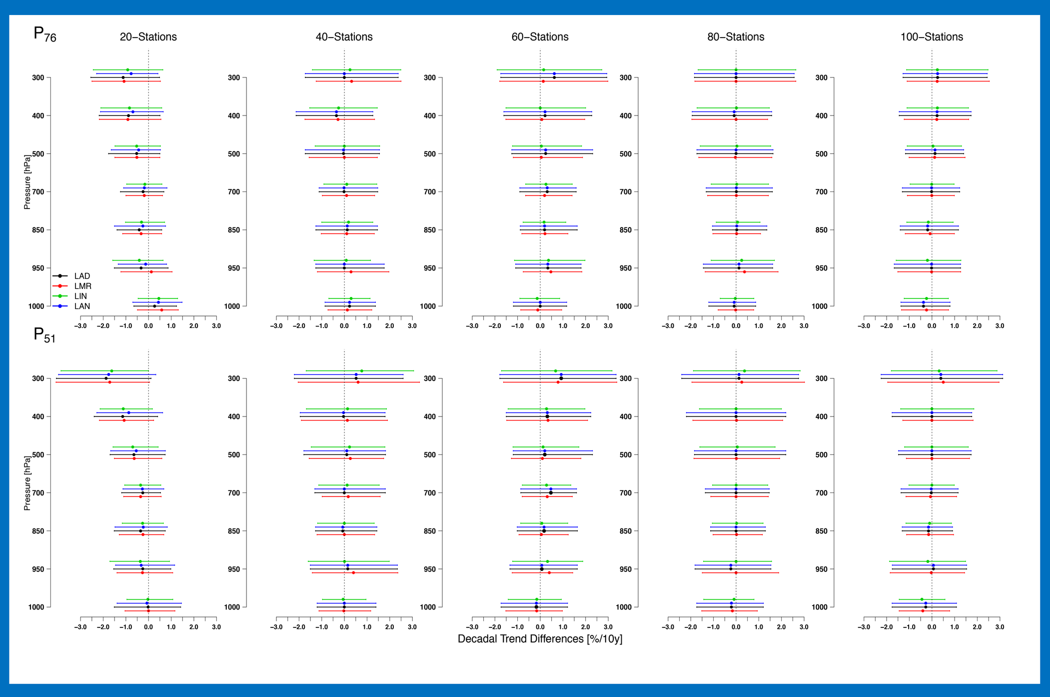

Figure 6. Same as Figure 5 but for relative humidity decadal Trends (%/decade).

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.

Susanti, Y., & Pratiwi, H. (2014). M estimation, S estimation, and MM estimation in robust regression. International Journal of Pure and Applied Mathematics, 91(3), 349-360.

4.Discussion & Conclusion:

- Trend differences can be influenced by several phenomena: i) sensitivity of each method to the specific nature of each dataset and to the presence of outliers; ii) quantity of datasets available in the time series (temporal sampling effects) and how it is incorporated by methods with reasonable goodness of fit and; iii) number of radiosonde stations selected to estimate trends (spatial sampling effects).
- Increasing the gaps of missing data in the time series of datasets can systematically increase the noise among the regression methods and this is effects are more evident in SH and tropics regions where radiosounding observations become limited. Subsampling uncertainties contributions to the uncertainty budget must be quantified to clearly demonstrate the value of any trend estimation, both in space and time and also in correlation with the selected trend estimation method for each specific application.
- These results highlight the need to always quantify the uncertainty contributions due to the choice of a regression estimation method and due to the effects of data subsampling affecting the time series in space and time.

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