

Quality Control and Homogenization of China's 6-hourly Station Pressure Data

Fang Yuan

National Meteorological Information Center, China



1. Introduction

2. Results

a. Quality Control

b. Elevation verification

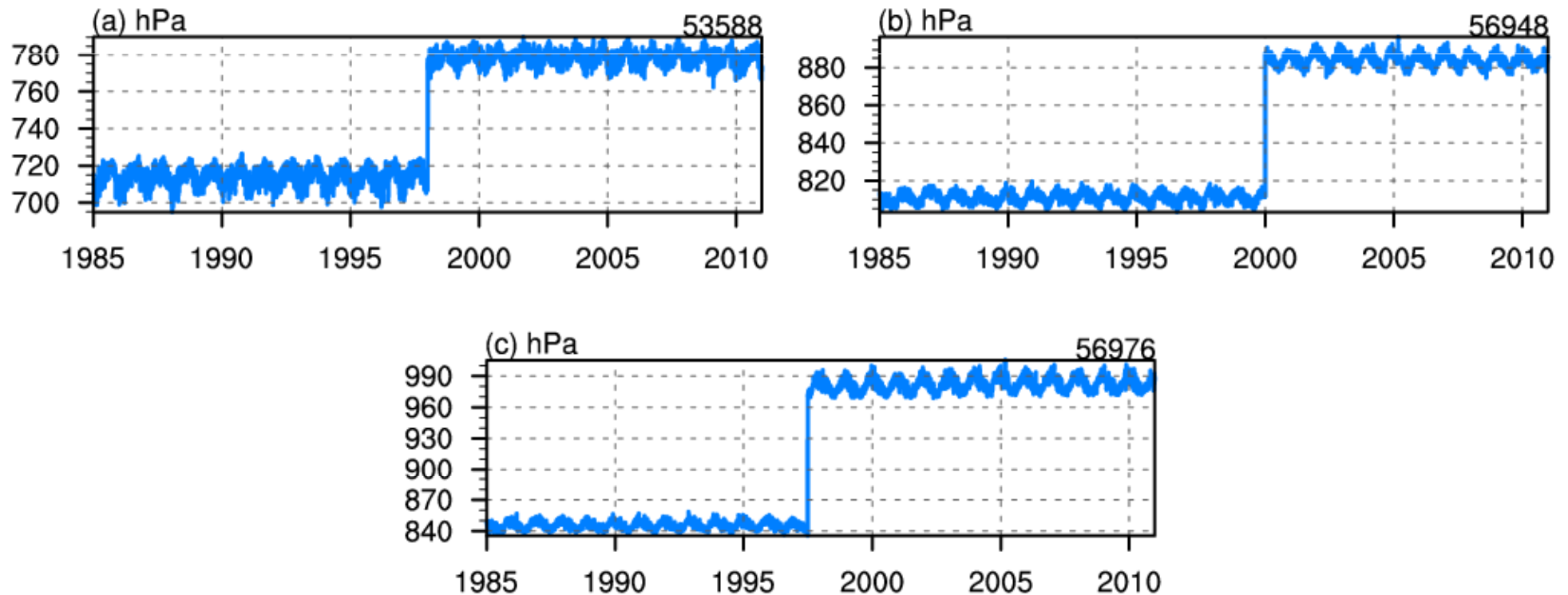
c. Homogenization

3. Summary



1. Introduction

Why we do it?



Station relocation is the most important factor to cause systematic error.



1.Introduction

Why SP, not sea level pressure(SLP)?

- The SLP data recorded in the archives in China is approximate data:

$$SLP=SP+C$$

$$C=34.68*h/(t+273)$$

h is elevation, and t is the annual mean temperature.



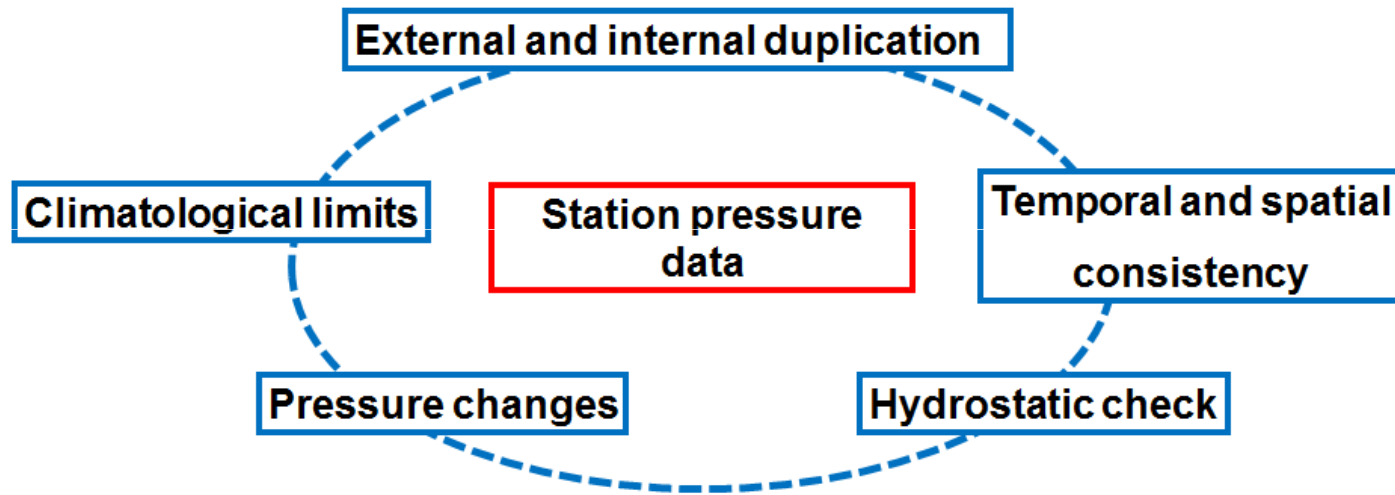
1.Introduction

How we do it?

- A comprehensive quality control (QC) system and a data homogenization procedure were applied to correct both random and systematic errors in 6-hourly station pressure (SP) data from **194** sites in China for the period 1951-2012.
- The penalized maximum F (**PMF**, Wang et al. 2008) test and quantile-matching (**QM**, Wang et al. 2010) algorithm were used for detecting and correcting systematic errors, separately.



2. Results - a. Quality control system



(1) Climatological limits

The **lowest and highest values in the 1971–2000 period** (for each station with at least 25 years of data in this period). These values must pass other QC checks.

(2) Pressure changes

For 12hr and 18hr pressure change, if over **3 σ** after standardization, the data will be flagged as suspicious.



(3) Hydrostatic check

Hydrostatic residuals R_z : $R_z = Z_m - Z$

$$Z_m = \ln \frac{P_0}{P_z} (T_0 + \overline{T_{\text{dry}}}) / \left(\frac{g}{R} - \frac{a}{2} \ln \frac{P_0}{P_z} \right)$$

- Z_m is the estimation of the station elevation obtained by substituting the related 6 hourly P_0 , P_z , or T_{dry} values.
- Z is the recorded current station elevation
- P_0 is SLP, P_z is station pressure
- $\overline{T_{\text{dry}}}$ is the average of the current dry-bulb temperature and the dry-bulb temperature recorded 12 h earlier ($^{\circ}\text{C}$).
- $T_0=273.15$, $a=0.0065^{\circ}\text{C m}^{-1}$, $R_d=287.085$, $g=9.8$

The tolerance of R_z is determined by sigma test

$$\mu - \gamma\sigma \leq R_z \leq \mu + \gamma\sigma$$

μ and σ are the mean and standard deviation of the R_z time series, γ is a parameter that defines the tolerance in terms of σ

Random error rate is fixed to **0.2‰** by dynamic γ for each station.



(4) Temporal and spatial consistency

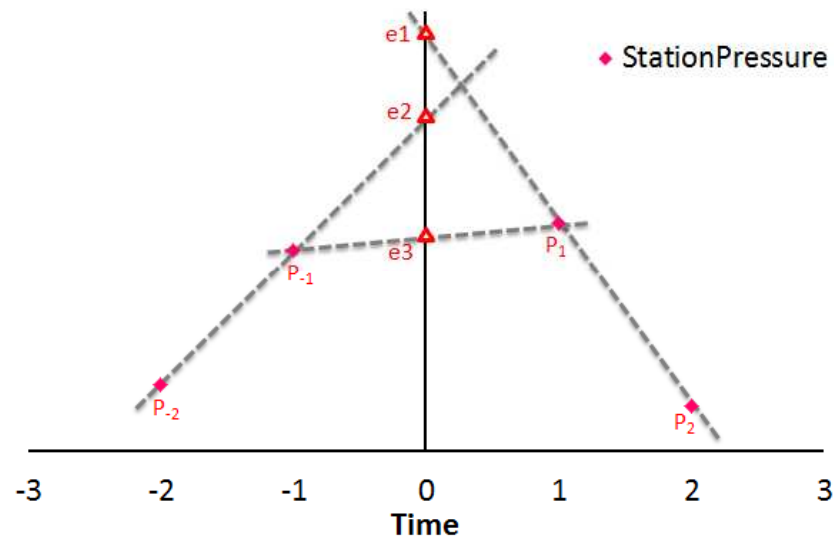
For each data of every station, a data will be flagged as suspicious if the difference with the data before and after it in time series or with the simultaneous data of it's nearest station is larger than 10 hPa.

(5) External and internal duplication

External duplication : Consecutive identical values of station pressure and SLP for more than 1 month.

Internal duplication: A constant value lasts over 3 days.

$$\begin{aligned}e1 &= 2 * P_1 - P_2 \\e2 &= 2 * P_{-1} - P_{-2} \\e3 &= (P_{-1} + P_1) / 2 \\P_0 &= (e1 + e2 + e3) / 3\end{aligned}$$



2. Results-b. Elevation verification

Example of station pressure data in the archives:

Sta	lat	lon	ele	Year	M	D	2000UTC	0000UTC	0800UTC	1400UTC
50527	4913	11945	6768	1955	12	30	9333	9350	9363	9389
50527	4913	11945	6768	1955	12	31	9402	9416	9431	9458
50527	4913	11945	6129	1956	1	1	9544	9562	9565	9573
50527	4913	11945	6129	1956	1	2	9562	9547	9509	9470
50527	4913	11945	6129	1956	1	3	9434	9421	9424	9426
50527	4913	11945	6129	1956	1	4	9405	9390	9404	9424
50527	4913	11945	6129	1956	1	5	9430	9440	9442	9454



2. Results - b. Elevation verification

Metadata

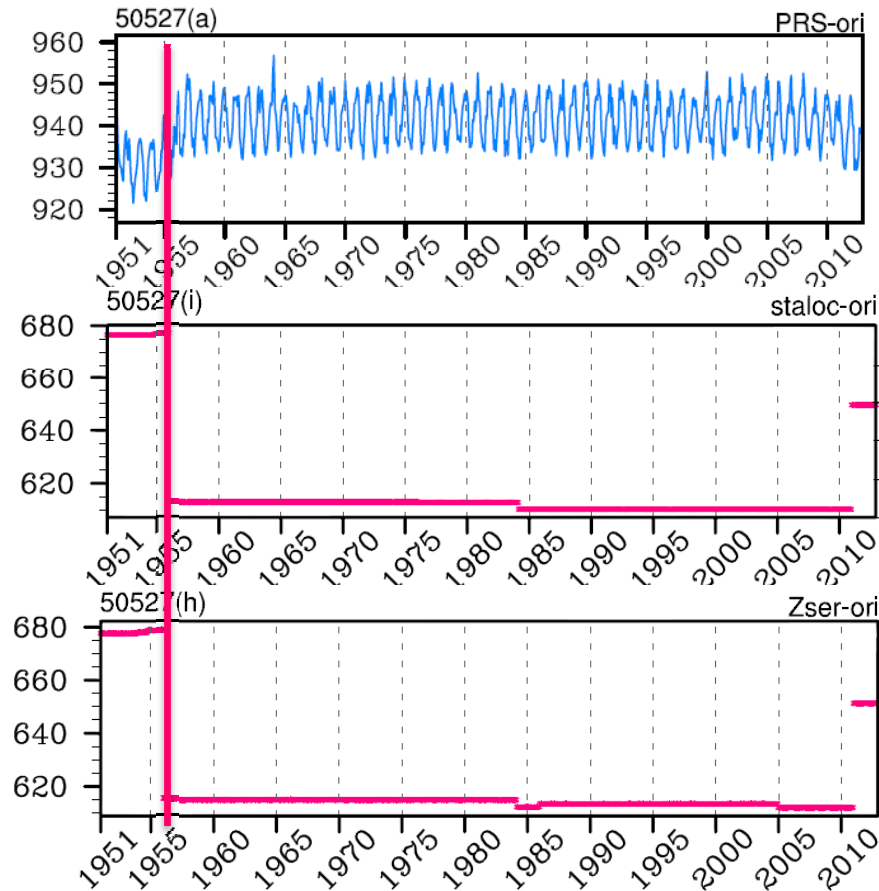
- 05 :relocation
- 55: re-measurement

Z_m series

- $Z_m = \ln \frac{P_0}{P_z} (T_0 + \overline{T_{dry}}) / \left(\frac{g}{R} - \frac{a}{2} \ln \frac{P_0}{P_z} \right)$



Example 150527:



original pressure data (hPa)

elevation recorded in archives (m)

Z_m series

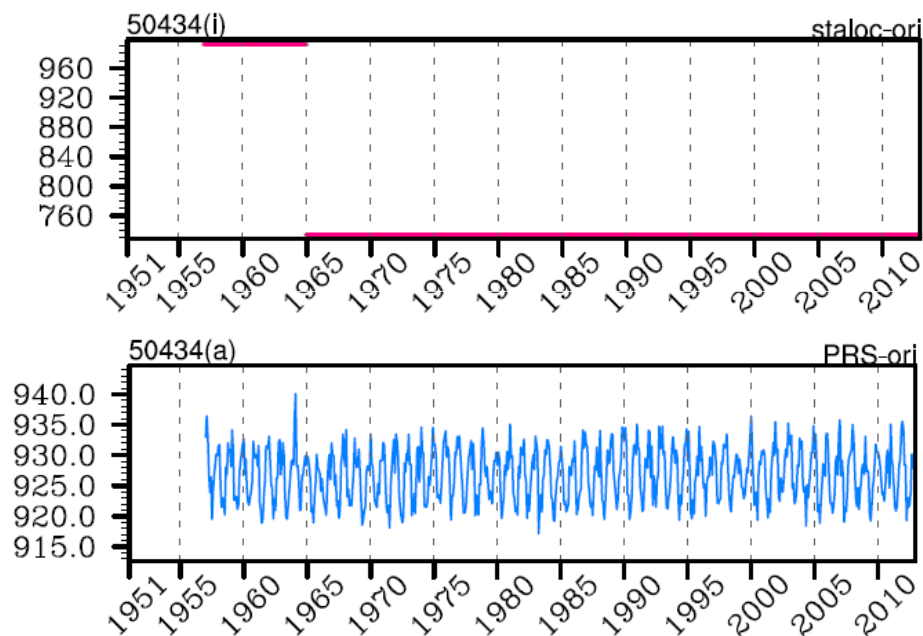
Metadata:

05/19500301/19530930/4913N/11945E/006766/
55/19531001/19551231/4913N/11945E/006768/
55/19560101/19601231/4913N/11945E/006129/

The metadata is wrong!



Example 250434:



elevation recorded in archives (m)

original pressure data (hPa)

Metadata:

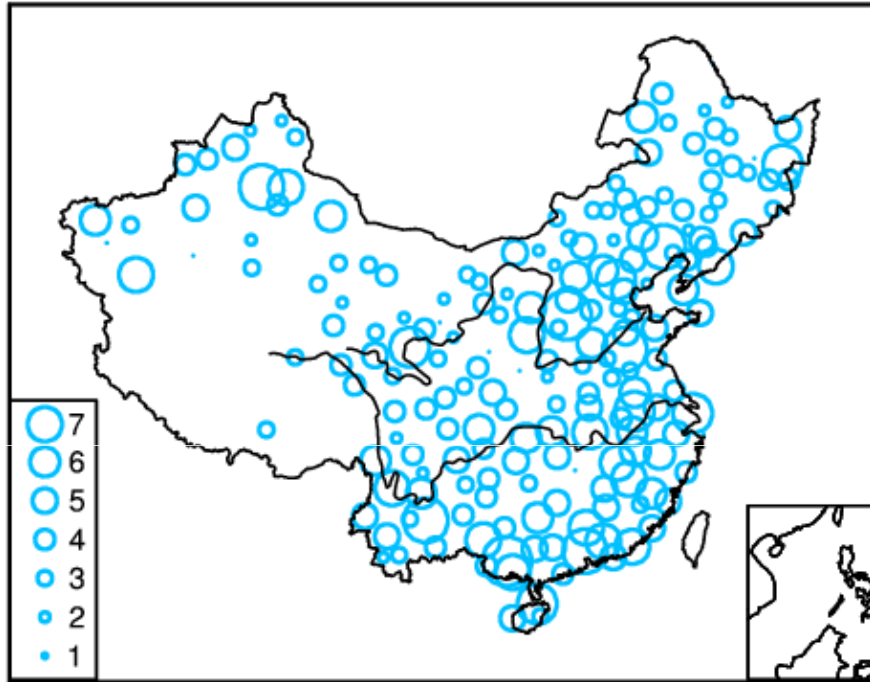
55/19610101/19641231/5030N/12128E/009936/

55/19650101/19661231/5030N/12128E/107325/

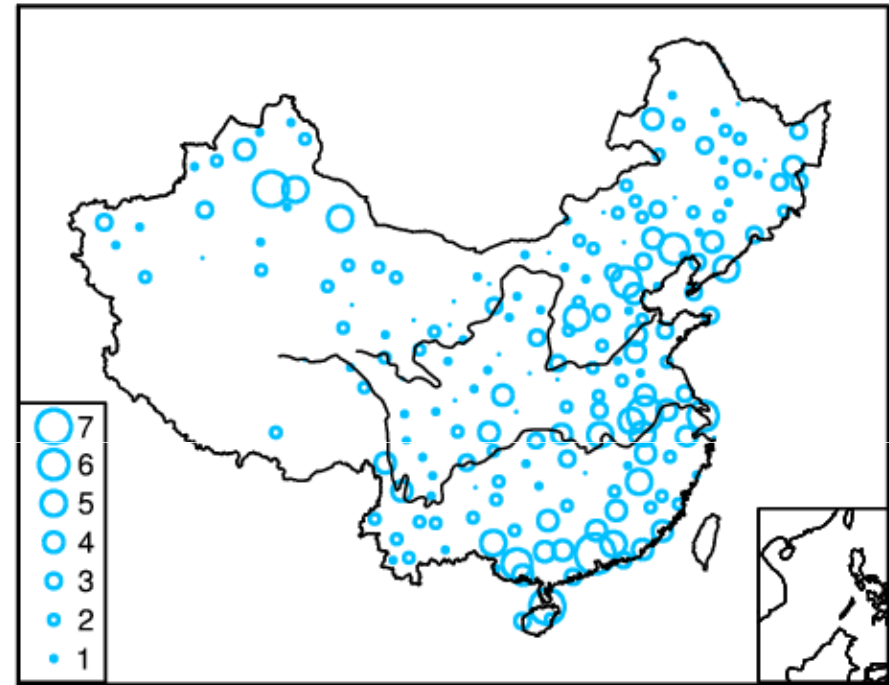
The elevation data is wrong!



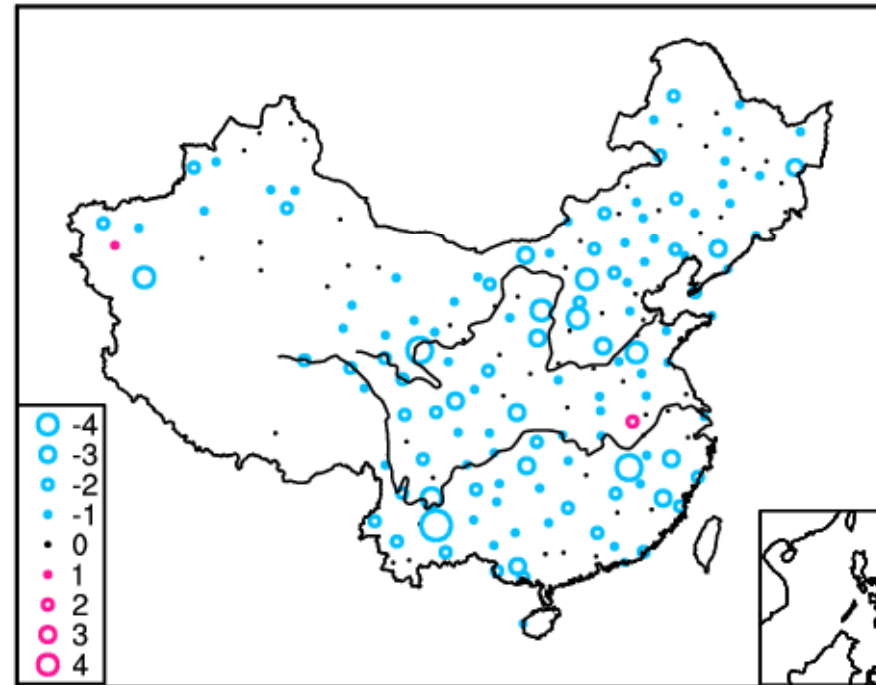
(a) Numbers of raw elevation changes



(b) Numbers of adjusted elevation changes



(c) difference

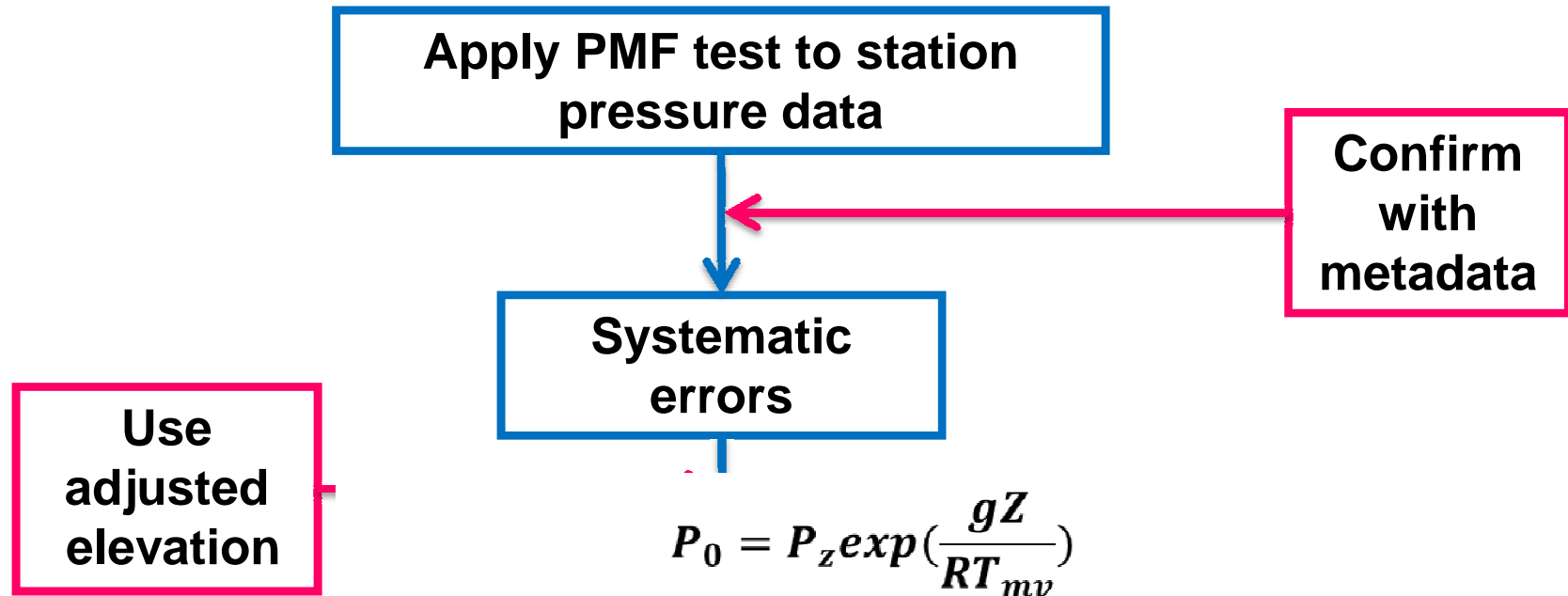


Numbers of elevation changes before (a) and after (b) confirmation and their difference (c).

$$\sum c = 214$$



2. Results-c. Homogenization



where

$$T_{mv} = (T_0 + \bar{T}_{dry}) + \frac{aZ}{2} + e_s C_h(Z) + F(\bar{T}_{dry})$$

$$e_s = (\bar{T}_{dry} + T_0)^{-0.00014\bar{T}_{dry}^2 + 0.0116\bar{T}_{dry} + 0.279}$$

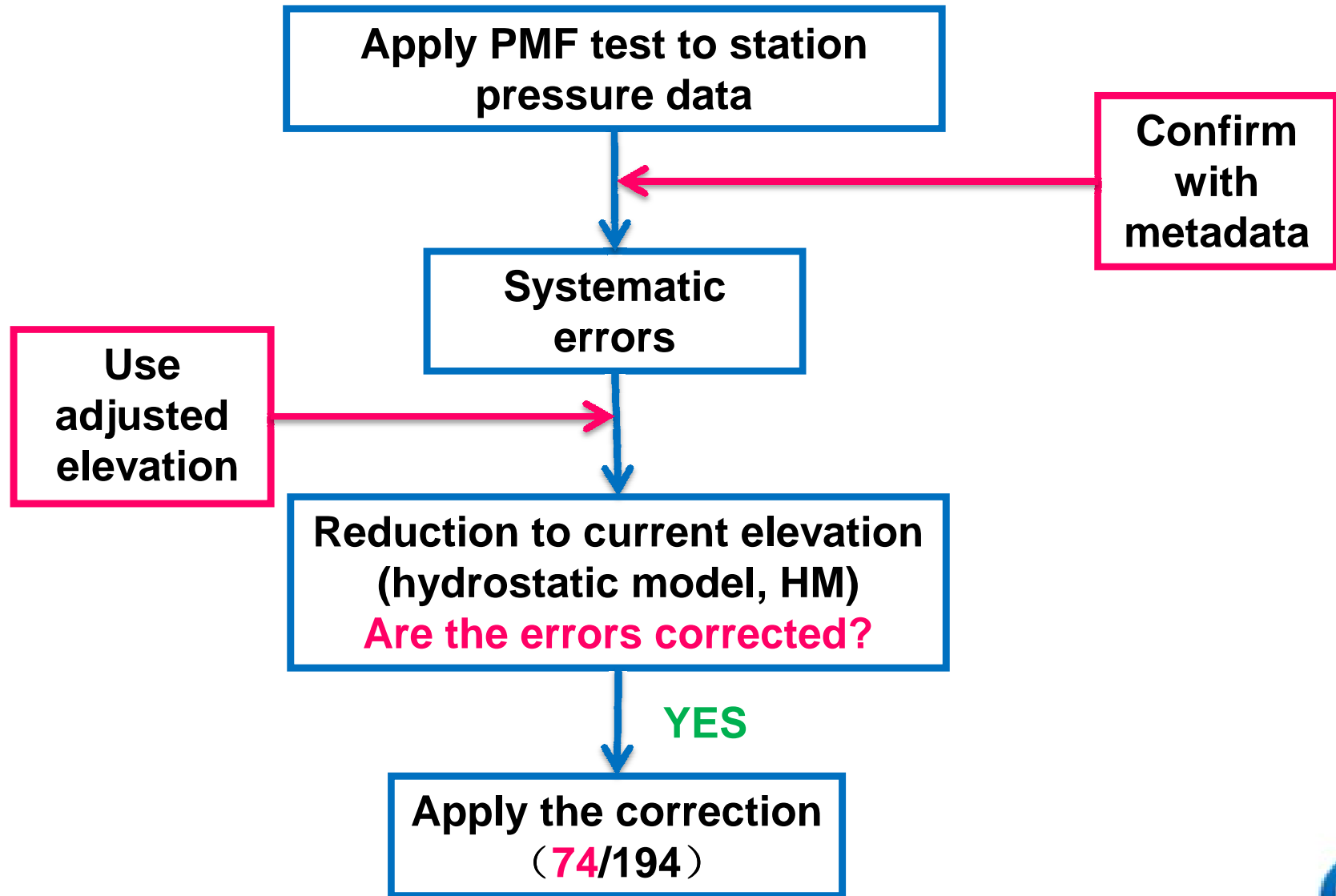
$$C_h(Z) = 2.8322 \times 10^{-9} Z^2 + 2.225 \times 10^{-5} Z + 0.10743$$

$$F(\bar{T}_{dry}) = b_1 \bar{T}_{dry}^2 + b_2 \bar{T}_{dry} + b_3$$

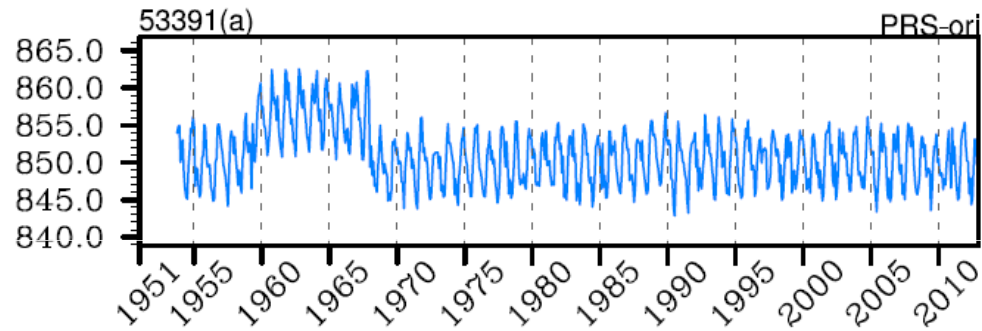
(Wan et al. 2007)



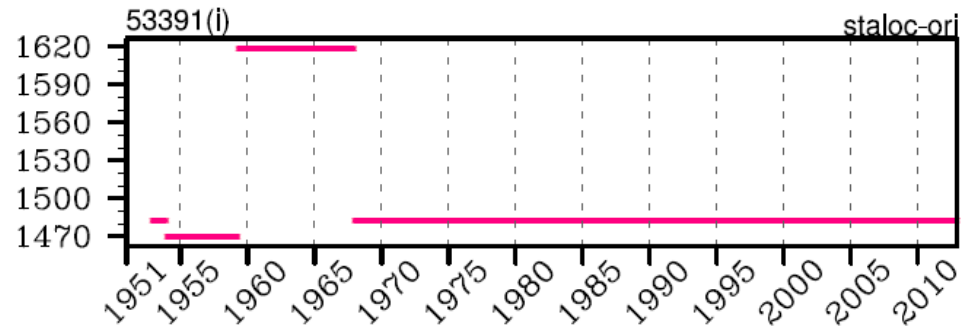
2. Results-c. Homogenization



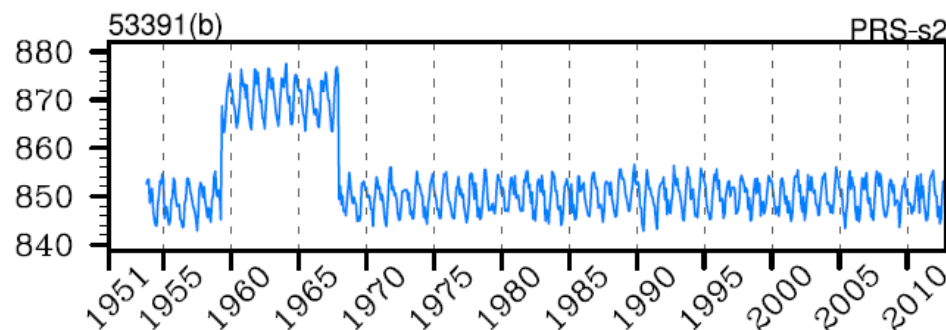
Example3 53391:



original pressure data (hPa)



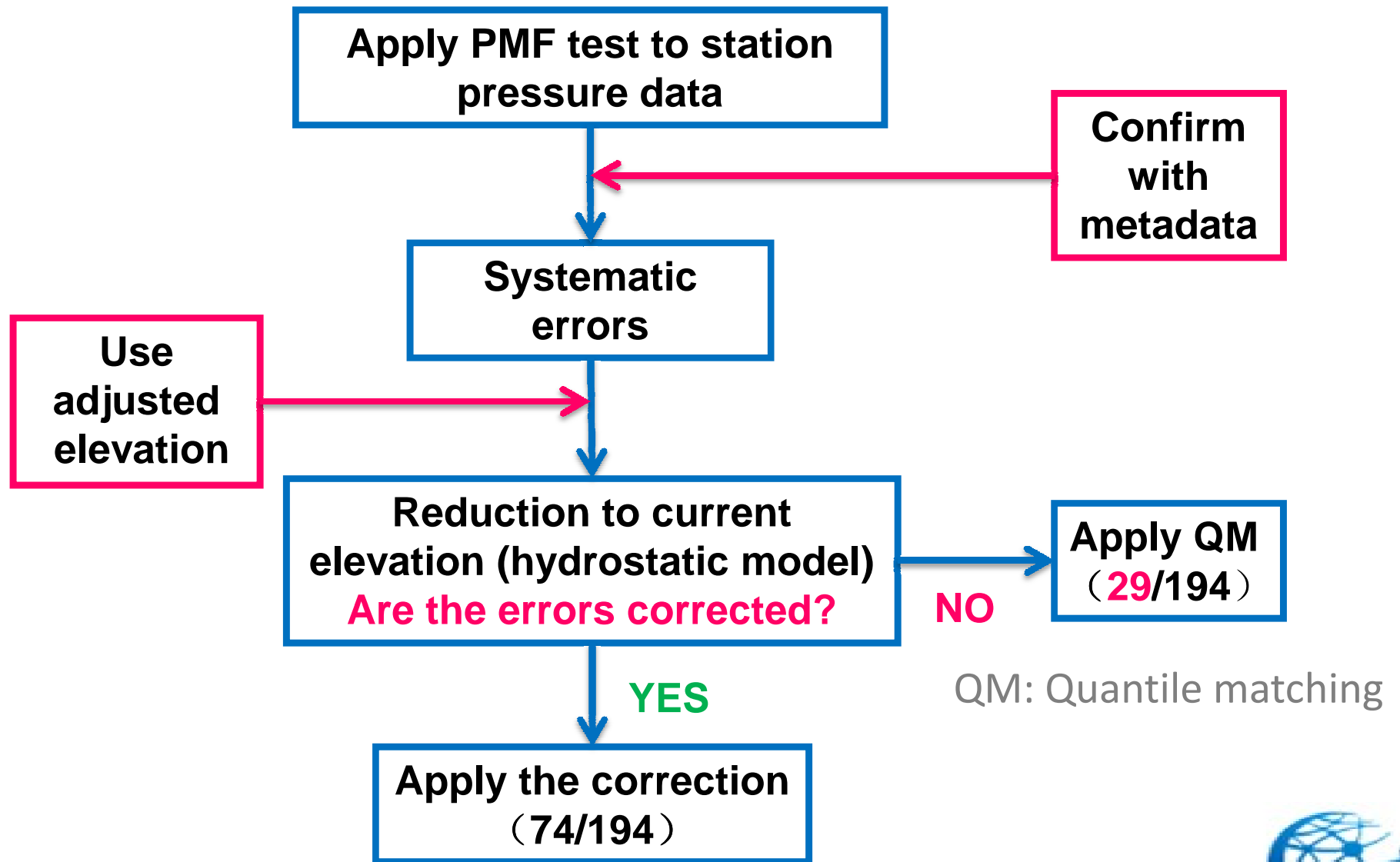
elevation recorded in archives
and also in metadata (m)



pressure data modified by
hydrostatic model (hPa)

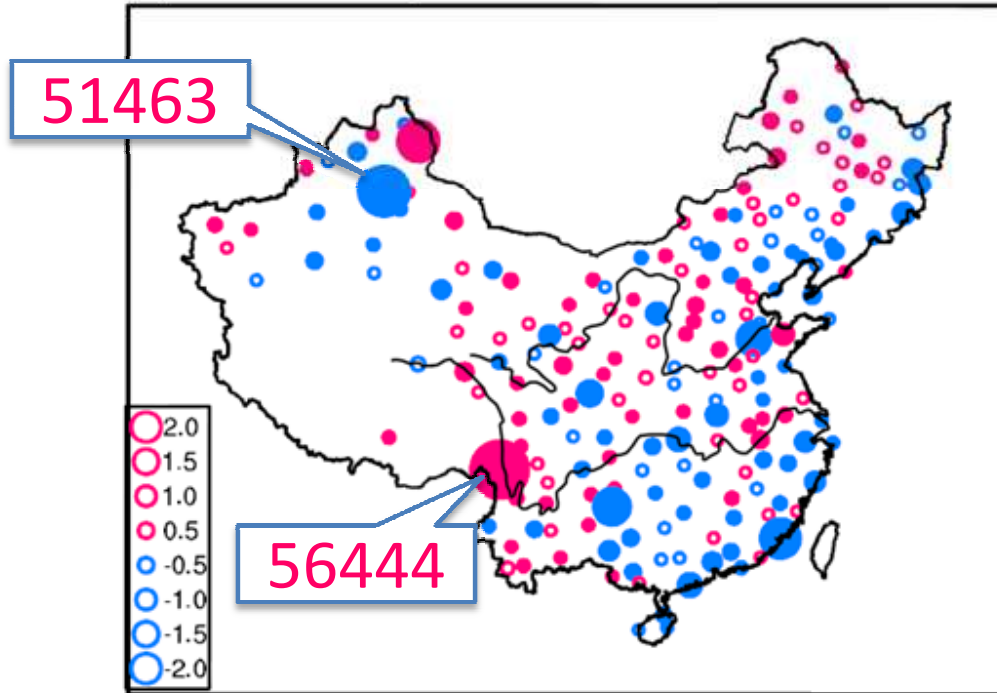


2. Results-c. Homogenization

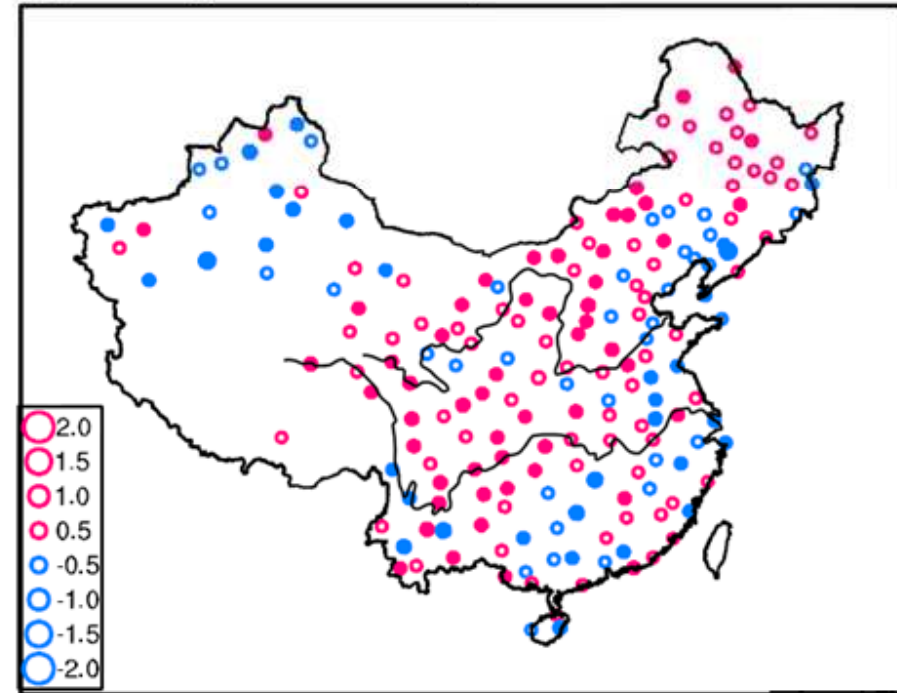


2. Results-Long term trend

(a) Raw station pressure trends



(b) Homogenized station pressure trends

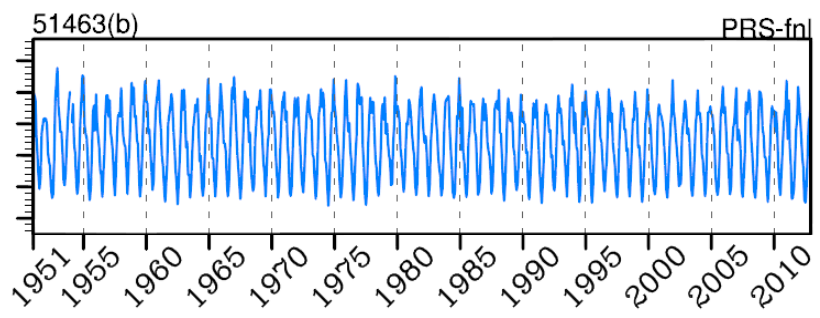
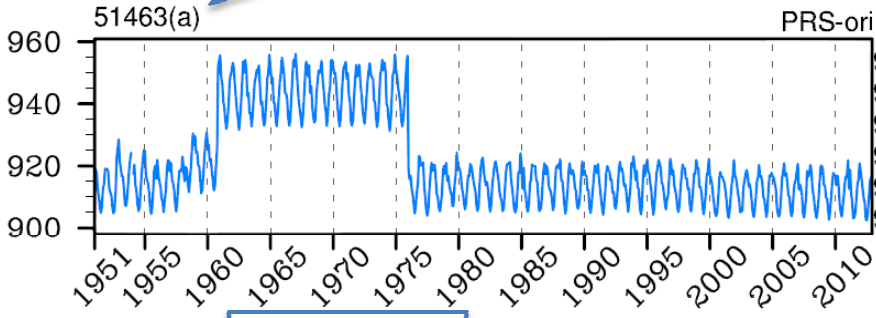


Long term trend of 194 stations (hPa/ decade): before (a) and after (b) modification. Filled circles indicate trends that are significant at the 5% level.

- Exclude extremely large values
- Better spatial consistency

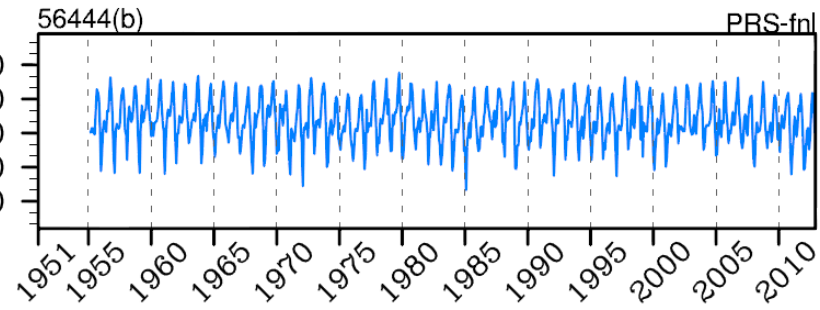
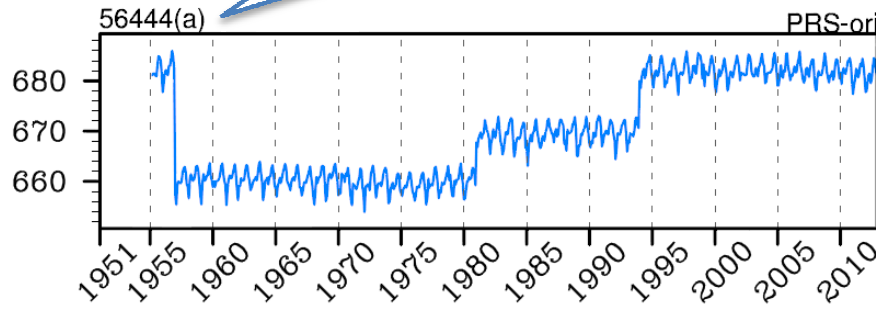


51463



HM

56444



QM

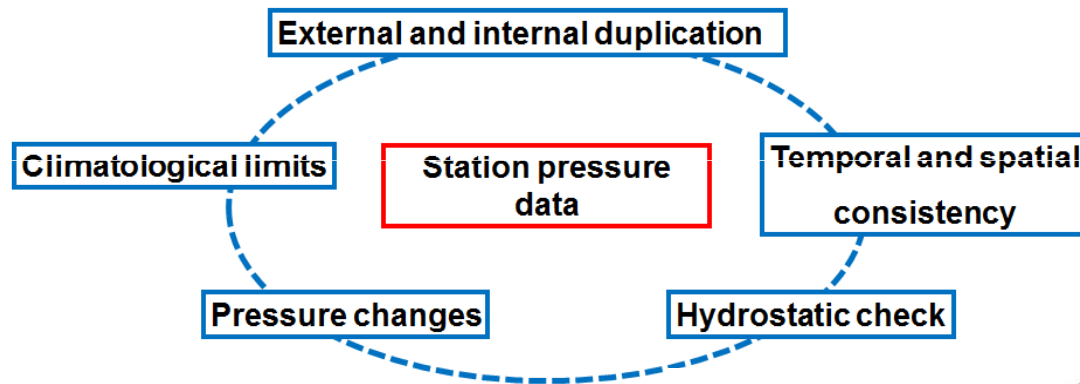
original pressure data

Modified pressure data

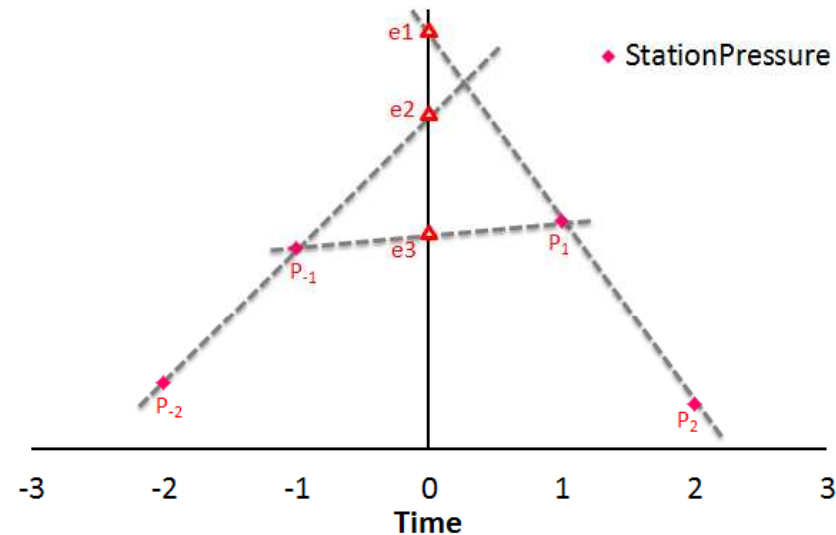
HM: Hydrostatic model
QM: Quantile matching



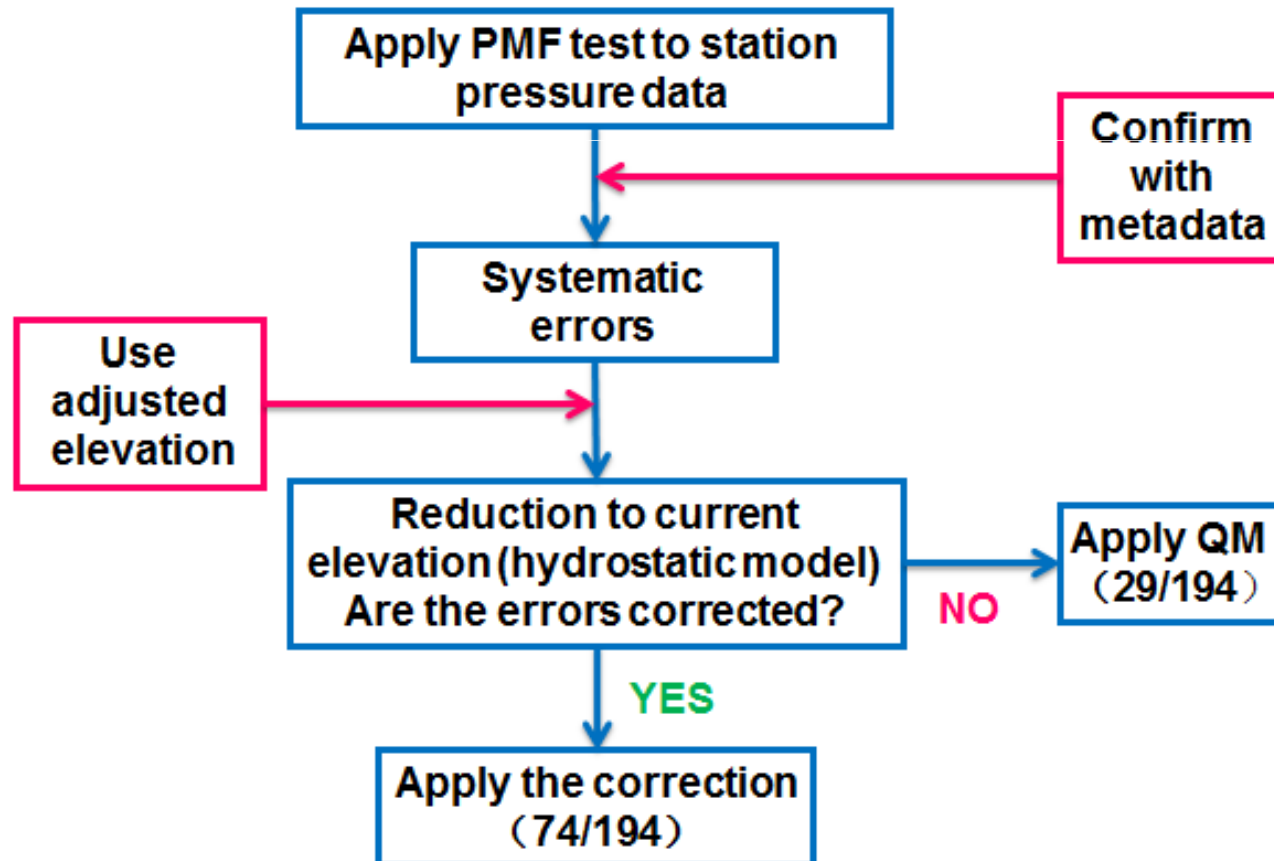
3. Summary - Quality control



$$e1 = 2 * P_{-1} - P_{-2}$$
$$e2 = 2 * P_{-1} - P_{-2}$$
$$e3 = (P_{-1} + P_{1}) / 2$$
$$P_0 = (e1 + e2 + e3) / 3$$



3. Summary - Homogenization



Thank you !

