A method of correction of climatological temperature records for the effect of relocation of meteorological stations.

Ohno, H., K. Yoshimatsu, K. Kobayashi, I. Wakayama, H. Morooka, Y. Oikawa, S. Hirahara, Y. Ikeda, H. Saitou
Presenter : S. Hagiya (clime@met.kishou.go.jp)
Japan Meteorological Agency, Tokyo, Japan

1.Introduction

- We developed a method of removing the influence from relocation of meteorological stations on monthly mean temperature data.
- The long term changes in temperature recorded at meteorological stations are important in understanding global and regional climate change.
- For stations which experienced relocation in the past, time-series data might not be homogeneous over a long period of time.
- In Japan, the climate of some station is different from the climate of nearby stations due to complexity of terrain of Japan.
- In the method of correction, we calculate a correction value for a relocated station by applying the principal component analysis (empirical orthogonal function analysis : EOF analysis) to monthly time-series data of the stations in Japan which have not been relocated.

2.Method

- We assume that a relocation of meteorological station is accompanied by a stepwise discontinuity in the temperature record.
- We express the monthly mean temperature of the station as Y in following multiple regression equation (1). In specifying the years we use the expression $Y(j)$, where j indicate year.

$$Y(j) = \underbrace{\sum_{l}^{L} a_l F_l(j)}_{(A)} + \underbrace{bS(j)}_{(B)} + \underbrace{e(j)}_{(C)} \qquad (1)$$

- **The term (A)** in Eq.(1) expresses the variations that is not due to relocation, where
  - F_l is the l -th EOF score that is calculated by applying the EOF analysis to monthly time-series data of the stations in Japan which have not been relocated.
  - L is the number of EOFs used for the analysis(the L major EOF's cumulative contribution ratio is 90%).
- **The term (B)** corresponds to the discontinuity due to relocation of stations. $S(j)$ is a step function defined by

$$\begin{aligned} S(j) &= -0.5 \quad \text{if} \quad j < j_h \quad \text{or} \quad j = j_h \;,\; m < m_k \\ S(j) &= +0.5 \quad \text{if} \quad j > j_h \quad \text{or} \quad j = j_h \;,\; m \geq m_k \end{aligned} \qquad (2)$$

where j_h and m_h is the year and month of the relocation of the station, respectively.
- **The term (C)** expresses the residual error.
  a_l and b in Eq.(1) are the least-squares coefficients which are determined on the condition of minimizing $e(j)$.
- The coefficient b is a correction value for a relocated station (℃). The corrected time-series data is obtained by adding the correction value b to the data before relocation.

3. Reference Period of the Data

- The reference period of monthly data used to calculate the correction value in Eq.(1) is determined to be 16 years (8 years preceding and succeeding the relocation). To determine the optimal reference period of the data,
  - We created artificial inhomogeneous time-series datasets by imposing a simulated stepwise discontinuity of five different values (±1℃, ±0.5℃, 0℃) at 1920, 1940, 1960 or 1980 of the time-series dataset of the station.
  - We calculate the correction value for the artificially created dataset. The optimal reference period of monthly data used to calculate was derived by extending the time window from 12 to 30 years by 2 year in order to estimate the accuracy of correction value.
- The accuracy of correction value becomes worse as extending the period of data in the time-series dataset of the station in urban area, and becomes better as extending the period of data in the time-series dataset of the station in rural area (Fig.1).
- The optimal reference period of the data is determined as 16 years to remove the discontinuity of time-series data of stations both in urban and rural area appropriately.

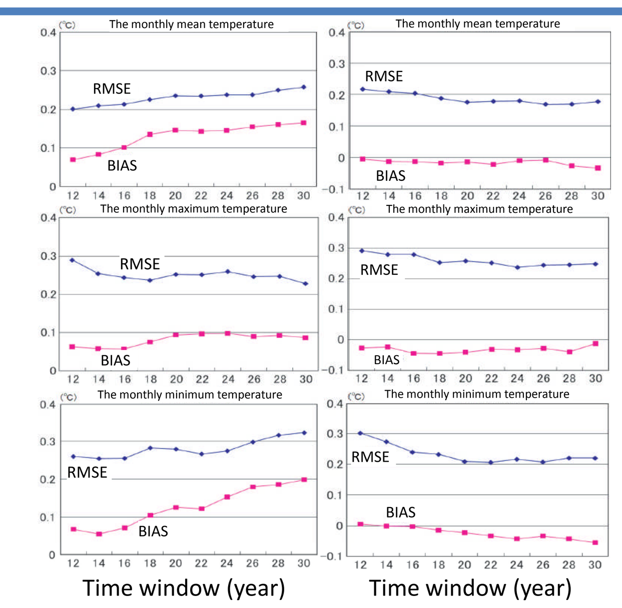

Fig.1 Performance dependency on time window (Left) BIAS (the difference between simulated step and calculated value, red line) and RMSE (root mean square of BIAS, blue line) averaged over stations in urban area (Tokyo, Osaka, Fukuoka(Japan)). (Right) same as Left but for stations in rural area (Suttsu, Iida, Hamada(Japan)).

4.Result of Correction

- We made a correction for 64 cases of the past relocation documented in Japan for meteorological stations that have time-series data in 50 years (25 years preceding and succeeding the relocation).
- We applied the two-phase linear regression test for changepoint detection (Wang(2003)) to the corrected dataset.
  The result of this test proved that, the ratio of datasets that are judged as not homogeneous has been improved from about 20% to 6%.
- 92% of the correction values we calculated (for monthly mean temperature, minimum temperature and maximum temperature of 64 relocations) are within the range of ±1℃.
- One of the correction value we calculated for a relocation in early 1900s was over estimated because of low station coverage in Japan.

- Example : Hiroshima(Japan)
  (moved about 2.5km to southwest in 1935, about 4.5km to northeast in 1988)

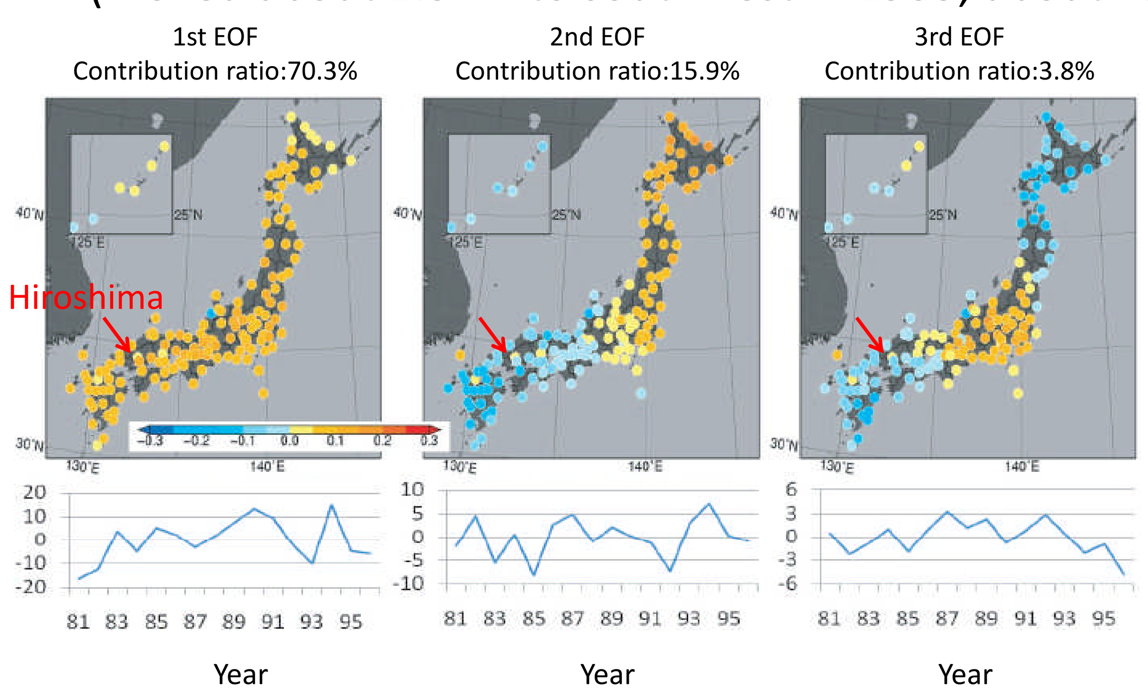

Fig.2 Major EOFs in Sep. 1988 for monthly mean temperature (Top) norm of EOF vector (normalized value, Degrees Celsius). (Bottom) EOF score (F_l in Eq.(1) (l =1-3), Degrees Celsius).

Year of relocation	1935			1988		
	T_{mean}	T_{max}	T_{min}	T_{mean}	T_{max}	T_{min}
Jan	0.0	−0.2	0.1	0.4	0.7	0.2
Feb	−0.4	−0.5	−0.4	0.6	0.7	0.4
Mar	−0.5	−0.7	−0.4	0.5	0.7	0.4
Apr	−0.5	−0.9	−0.3	1.1	1.4	0.8
May	−0.9	−1.4	−0.7	1.1	1.4	0.5
Jun	−0.9	−1.7	−0.6	0.7	1.3	0.4
Jul	−1.0	−1.4	−0.7	1.5	1.5	0.8
Sug	−1.0	−1.6	−0.9	0.9	1.2	0.6
Sep	−0.8	−1.4	−0.8	1.0	1.6	0.6
Oct	−0.5	−0.6	−0.6	0.2	0.9	0.2
Nov	−0.3	−0.1	−0.4	0.3	0.5	0.1
Dec	−0.1	−0.1	−0.2	0.4	0.5	0.0

Table.1 The list of correction value of Hiroshima (degree Celsius)
The corrected time-series dataset is obtained by adding the correction value to the data before relocation.
T_{mean} , T_{max} and T_{min} correspond to the monthly mean, maximum, minimum temperatures, respectively.

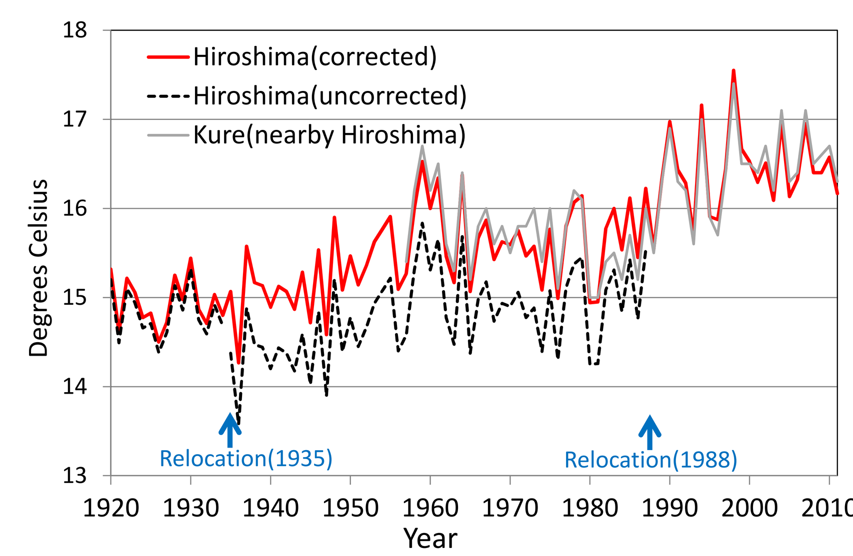

Fig.3 Corrected yearly mean temperature (red solid line) and uncorrected yearly mean temperature (black dashed line) of Hiroshima
Gray solid line correspond to yearly mean temperature of Kure(the station nearby Hiroshima).

5.Summary

- In the homogenization method presented here, time-series temperature records are adjusted for the discontinuity arising from station relocations using the adjustment values and natural annual variabilities, which are estimated by applying principal component analysis constructed from monthly mean temperature series from domestic stations that have not been influenced by relocation.
- The two-phase linear regression test for changepoint detection (Wang(2003)) concludes that only 6% of the 64 adjusted time-series datasets are judged to be inhomogeneous, demonstrating an improvement from 20% of the unadjusted ones.
- If station coverage in Japan is low, the variation of temperature record may not be reproduced sufficiently.
- We should estimate the impact of non climatic variations due to man-made causes comprehensively.

References : Ohno et al. (2011), Sokkou Jihou, 78. 31-41 (in Japanese)
Wang, X. L. (2003): Comments on "Detection of undocumented changepoint: A revision of the two-phase regression model." J.Climate, 16, 3383-3385