ATLANTIS PRESS

Advances in Engineering Research (AER), volume 143

6th International Conference on Energy and Environmental Protection (ICEEP 2017)

Quality Control for Surface Hourly Temperature Observations via An Improved SRT Method

ZHANG Qidong[1, a] and XIONG Xiong[2,b]

[1]State Grid Jiangsu Electric Power Company Maintenance Branch, Nanjing 211102, China

[2]School of Information and Control, Nanjing University of Information Science and Techology, Nanjing 210044, China

[a]oicq_5@163.com, [b]nxgxiong@163.com

Keywords: surface hourly temperature, quality control, RH-SRT, RH, SRT.

Abstract. The article aims to discuss the quality control methods for the surface hourly temperature observations. A new quality control method (RH-SRT) based on the spatial regression test (SRT) was adopted to identify potential outliers in the surface hourly temperature observations. The proposed method is an extension of SRT, which takes the relationship between relative humidity and temperature as the explaining variable to obtain the estimated value of surface temperature. In order to evaluate the proposed method, RH-SRT was applied to an annual temperature dataset with seeded errors for different regions in Jiangsu Province during 2007. Compared with the traditional SRT method, the results demonstrate that RH-SRT outperforms SRT for all cases. Moreover, comparison results demonstrate that the RH-SRT method is an effective version of SRT, which could be used to improve the quality of surface hourly temperature observations.

Introduction

A continually increasing number of meteorological observation sites is producing larger and larger amounts of surface data [1]. Surface data is essential to our effort to identify and understand variations of regional and global climate [2]. First, high quality long-term observations are necessary for identifying climate changes or for validating climate model simulations [1]. Moreover, quality controlled real-time data are fundamental for new casting and model validation and furthermore are used to provide proper initial conditions for numerical weather prediction [3]. However, the surface observations are easily affected by the stations positions, observing instruments, and human factors and so on [4-7]. Thus, it is important to carry out quality control (QC) before surface observations applications [2, 8, 9].

Traditional QC methods can be divided into two principal categories: the first one is designed for a single station and the other relies on the using of multiple stations [4, 10-13]. For a single station, QC methods such as extreme check, internal consistency check, temporal outliers check and spatial outliers check have been used in seeking out the potential outliers. In recent years, QC methods based on multiple stations have been found to be effective, especially when extreme events occur. The key idea of these methods is estimating the value of target station against the observations from neighboring stations. In 2005, Hubbard [14] and You [15] proposed a spatial regression QC method (SRT) to identify the outliers in the surface observations. SRT calculates the estimates of the target station according to the standard error between the target station and neighboring stations, which had been proved outperforming the IDW.

In order to improve the stability of the QC method, a complex QC approach (RH-SRT) is employed herein to identify the seeded errors in the surface hourly temperature observations.

The RH-SRT method

SRT [4] is a quality control method that checks out the potential error data according to the neighboring station data during a time period n. The neighboring stations are selected within a

certain distance of the target station. For each neighboring station paired with the target station, a linear regression based on estimate is formed:

$$x_i = a_i + b_i \cdot y_i \tag{1}$$

where y_i is the data for the ith neighboring station $(i = 1,2,3\mathrm{L}\ ,n)$, a_i and b_i are the regression coefficients, x_i is the estimate of the target station based on y_i. The weighted estimate $\overline{x}$ is obtained by using the standard error of estimate s:

$$\overline{x} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 \cdot s_i^{-2}}{\sum_{i=1}^{n} s_i^{-2}}} \tag{2}$$

The weighted standard error of estimate $\overline{x}$ is calculated from:

$$\frac{1}{\overline{x}^2} = \frac{1}{n} \cdot \sum_{i=1}^{n} s_i^{-2} \tag{3}$$

Generally, RH can be defined as the ratio of the actual to the saturation vapor pressure:

$$RH = \frac{e_a}{e_s} \times 100\% \tag{4}$$

where e_a and e_s are the actual vapor pressure and the saturation vapor pressure, respectively. e_s can be defined as a unique function of temperature. In order to depict the change of e_s accurately, we calculate the definite integrals of the change of RH to reconstruct RH. Here, a new formula for estimation of e_s from temperature based on Gaussian Model (Gaussian formula) is defined as:

$$e_s = a \cdot e^{-(b \cdot T + c)^2} \tag{5}$$

where a, b and c are undetermined constants, T is the temperature in degrees Celsius. we obtain that:

$$T = 151.4 - 64.9 \cdot \sqrt{\ln(1397 \cdot RH) - \ln(e_a)} \tag{6}$$

Equation. (10) can be employed to calculate the temperature according the relative humidity and the actual vapor pressure.

In this work, we proposed an efficient and convenient quality control method (RH-SRT) for surface hourly air temperature. RH-SRT consists of two parts: the first one is based on the relationship between temperature and relative humidity, the second one is based on SRT. The RH-SRT can be defined as:

$$T_{est} = a \cdot T_{RH} + (1 - a) \cdot T_{SRT} \tag{7}$$

$$a = \frac{r_{RT}}{r_{RH} + r_{SRT}} \tag{8}$$

where T_{est} is the estimate value of the target station, T_{RH} is the estimate value of the target station T_{RH} and T_{SRT} are based on Eq. (6) and Eq. (1), respectively. r_{RH} is the correlation coefficient between T_{RH} and observations of the target station for a nearest period, and r_{SRT} is the correlation coefficient between T_{SRT} and observations of the target station.

The confidence intervals are given by the standard error s of the target stations:

$$\left|T_{est} - T_{obs}\right| \le f \cdot s \tag{9}$$

where T_{obs} is the observation value of the target station, f is the threshold. If the observations of the target station fall within the confidence intervals, the observations would pass the RH-SRT test.

Results and discussion

In the test of RH-SRT, we applied RH-SRT to the dataset with seeded errors and compared it with SRT. A climate dataset based on four regions (Nanjing (NJ), Wuxi (WX), Xuzhou (XZH) and Lianyungang (LYG)) in Jiangsu Province for year 2007 was used to validate the method. The Nash--Sutcliffe model efficiency coefficient (NSC), mean absolute error (MAE), root-mean-square error (RMSE), coefficient of determination (R^2) and MSR are used here as a measure to evaluate the new method.

An example of the analysis for both two methods for sensitivity to radius, number of stations, window length and offset is presented in Fig. 1. As shown in Fig. 1a, when the radius is greater than 90 km, MAE for different regions is relatively stable or only decline slightly. When the radius is less than 90 km, however, some stations have large fluctuations. Another example is used to evaluate the performance of RH-SRT method according to different number of stations (see Fig. 1b), the reference stations are selected by the distance to the target station. It is illustrated that 15 stations are recommended for the different regions according to the figure. The ratio of detected seeded errors to the total number of seeds oscillates for different windows. Fig. 1c demonstrates that the RH-SRT method has the best ratio when the window length is 7 days. The performances of RH-SRT changes slightly with the different window offsets, but it is also obvious that the ratios are the best when the window is centered on the date (see Fig. 1d). In this study, we suggest starting a window length equal to 7 days for RH-SRT, using the centered offset and adopting 90 km for the radius for best results.

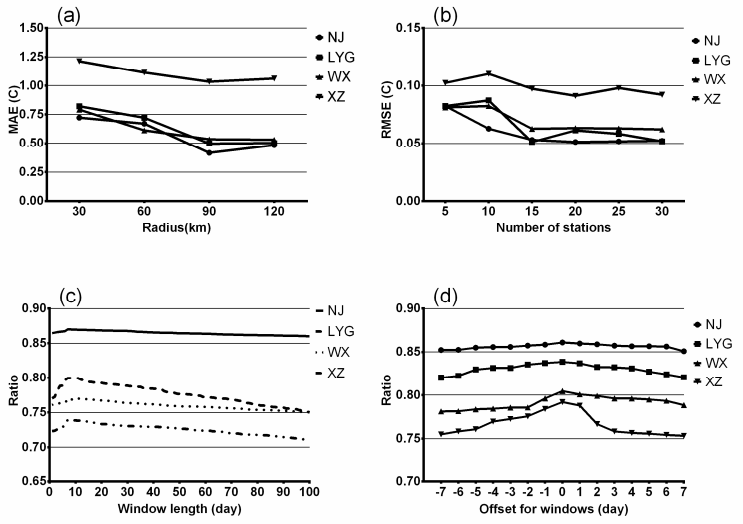

Fig.1 Examples of performance of RH-SRT and SRT with regard to radius (a), number of station (b), window length (c) and offset (d) for four different regions: Nanjing(NJ), Lianyungang(LYG), Wuxi(WX) and Xuzhou(XZ).

Figure 2 compares the distribution of MSR, MAE, RMSE and NSC for the RH-SRT method and the SRT method for the each month in 2007 individually. It demonstrates that the RH-SRT does better than SRT, especially in Figures 2b and 2d. The distributions almost have the similar tendency for these two methods except in July and August. A more detailed analysis can be seen in Figure 3, which illustrates that RH-SRT has a better stability than SRT, the standard deviation (SD) of the scatters for RH-SRT is 0.9237 while 0.7841 for SRT. The distribution of the MSR scatters for RH-SRT is closer and gathered around the CD line than SRT, especially in July and August. A possible reason is that there are more extreme weather occurred in July and August in Jiangsu Province. However, even in summer, RH-SRT has a stable and effective performance.

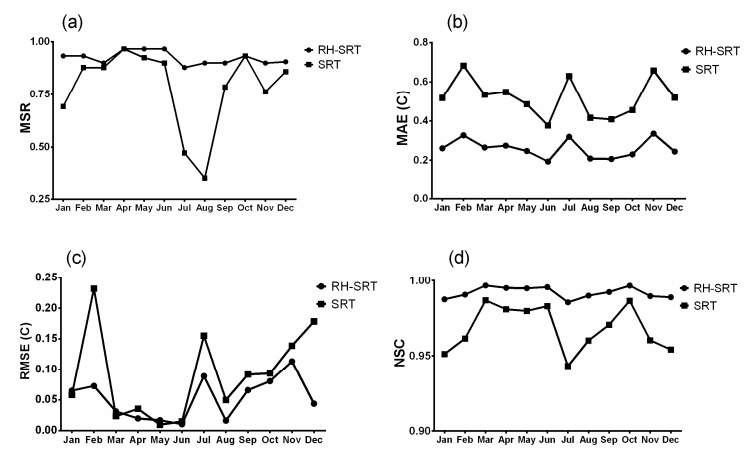

Fig. 2 The MSR(a), MAE(b), RMSE(c) and NSC(d) for RH-SRT and SRT in Nanjing region for each month in 2007.

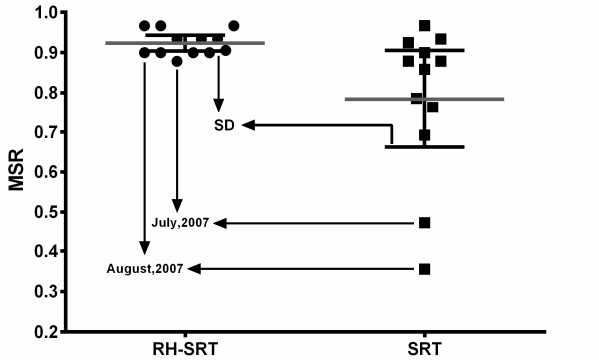

Fig. 3 The MSR with 95 percent confidence interval in Nanjing region for the year 2007 for both the RH-SRT method and the SRT method.

Figure 4 illustrates the regression between observed and estimated values for these four regions in Jul, 2007. Using the proposed method, the slopes of the regression lines are better than SRT for all cases. As shown in Figs. 4a-h, R^2 obtained by RH-SRT is better than that obtained by SRT, especially in Xuzhou (see Figs. 4g-h). The same results can also been seen from the performance of RMSE, the RMSEs for RH-SRT are significantly better than others for SRT. A possible explanation for these is that RH-SRT has the advantages of spatial check and temporal check.

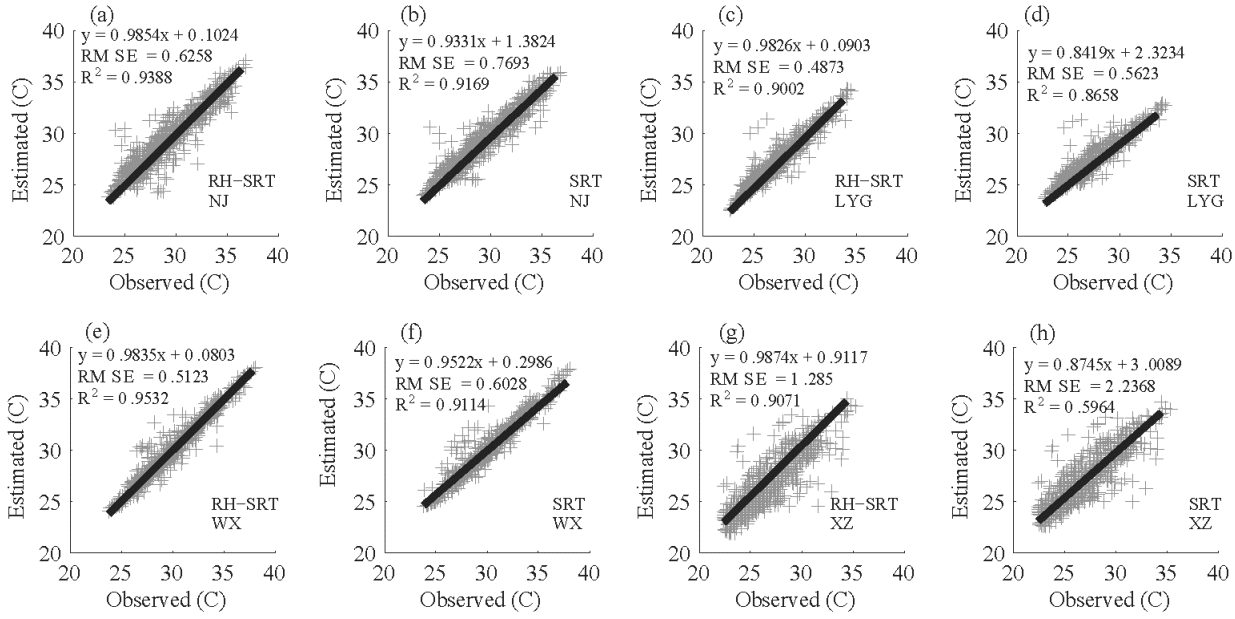

Fig. 4 Comparison of observations to estimates for RH-SRT and SRT for different regions (NJ, LYG, XZ, WX) for surface hourly temperature for July in 2007.

Conclusions

A new quality control method (RH-SRT) was proposed for identifying the potential outliers in surface hourly temperature observations. The proposed method utilizes the relationship between temperature and relative humidity and SRT to obtain the estimates of the target station. In order to evaluate the new method, SRT, as the standard and baseline method, is employed to compare the performance of RH-SRT. The RH-SRT method was found to be significantly better than SRT for all cases. This is reasonable because RH-SRT is an extension of SRT. Actually, RH-SRT takes the relationship between relative humidity and temperature as an explaining variable in quality control for the surface hourly temperature observations. Meanwhile, a new formula was presented to reconstruct the temperature via the relative humidity and saturation vapor pressure.

In general, increasing f decreases the probability of type I error but increases the probability of type II error. The balance between type I error and type II error is essential in the quality control procedures for the surface hourly temperature. In this manuscript, we present a new measure (MSR) to balance these two type of errors. Overall, MSR can be successfully used to select the best value of f for the quality control procedures.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (Grant No. 41675156).

References

[1] Dj.M. Maric, P.F. Meier and S.K. Estreicher: Mater. Sci. Forum Vol. 83-87 (1992), p. 119

[2] M.A. Green: *High Efficiency Silicon Solar Cells* (Trans Tech Publications, Switzerland 1987).

[3] Y. Mishing, in: *Diffusion Processes in Advanced Technological Materials*, edtied by D. Gupta Noyes Publications/William Andrew Publising, Norwich, NY (2004), in press.

[4] G. Henkelman, G.Johannesson and H. Jónsson, in: Theoretical Methods in Condencsed Phase Chemistry, edited by S.D. Schwartz, volume 5 of Progress in Theoretical Chemistry and Physics, chapter, 10, Kluwer Academic Publishers (2000).

[5] R.J. Ong, J.T. Dawley and P.G. Clem: submitted to Journal of Materials Research (2003)

[6] P.G. Clem, M. Rodriguez, J.A. Voigt and C.S. Ashley, U.S. Patent 6,231,666. (2001)

[7] Information on http://www.weld.labs.gov.cn

[1] Steinacker, R., D. Mayer, and A. Steiner, Data quality control based on self-consistency. Monthly Weather Review, 2011. 139(12): p. 3974-3991.

[2] Feng, S., Q. Hu, and W. Qian, Quality control of daily meteorological data in China, 1951–2000: a new dataset. International Journal of Climatology, 2004. 24(7): p. 853-870.

[3] Ingleby, N.B. and A.C. Lorenc, Bayesian quality control using multivariate normal distributions. Quarterly Journal of the Royal Meteorological Society, 1993. 119(513): p. 1195-1225.

[4] Hubbard, K.G. and J. You, Sensitivity analysis of quality assurance using the spatial regression approach-A case study of the maximum/minimum air temperature. Journal of Atmospheric and Oceanic Technology, 2005. 22(10): p. 1520-1530.

[5] Zhao, H., X. Zou, and Z. Qin, Quality control of specific humidity from surface stations based on EOF and FFT—Case study. Frontiers of Earth Science, 2015. 9(3): p. 381-393.

[6] Xu, Z., Y. Wang, and G. Fan, A Two-Stage Quality Control Method for 2-m Temperature Observations Using Biweight Means and a Progressive EOF Analysis. Monthly Weather Review, 2013. 141(2): p. 798-808.

[7] Xu, C.-D., et al., Estimation of Uncertainty in Temperature Observations Made at Meteorological Stations Using a Probabilistic Spatiotemporal Approach*. Journal of Applied Meteorology and Climatology, 2014. 53(6): p. 1538-1546.

[8] Alexander, L., et al., Global observed changes in daily climate extremes of temperature and precipitation. Journal of Geophysical Research: Atmospheres (1984–2012), 2006. 111(D5).

[9] Steinacker, R., D. Mayer, and A. Steiner, Data Quality Control Based on Self-Consistency. Monthly Weather Review, 2011. 139(12).

[10] Kubecka, P., A possible world record maximum natural ground surface temperature. Weather, 2001. 56(7): p. 218-221.

[11] Gleason, Global daily climatology network. National Climatic Data Center, 2002. V 1.0.

[12] Meek, D. and J. Hatfield, Data quality checking for single station meteorological databases. Agricultural and Forest Meteorology, 1994. 69(1): p. 85-109.

[13] Lanzante, J.R., 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, 1996. 16(11): p. 1197-1226.

[14] Hubbard, K.G., et al., An improved QC process for temperature in the daily cooperative weather observations. Journal of Atmospheric and Oceanic Technology, 2007. 24(2): p. 206-213.

[15] You, J. and K.G. Hubbard, Quality control of weather data during extreme events. Journal of Atmospheric and Oceanic Technology, 2006. 23(2): p. 184-197.