Estimation of Uncertainty in Temperature Observations Made at Meteorological Stations Using a Probabilistic Spatiotemporal Approach*

CHENG-DONG XU, JIN-FENG WANG, AND MAO-GUI HU

State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

QING-XIANG LI

National Meteorological Information Center, Beijing, China

(Manuscript received 23 May 2013, in final form 30 October 2013)

ABSTRACT

A probabilistic spatiotemporal approach based on a spatial regression test (SRT-PS) is proposed for the quality control of climate data. It provides a quantitative probability that represents the uncertainty in each temperature observation. The assumption of SRT-PS is that there might be large uncertainty in the station record if there is a large residual difference between the record estimated in the spatial regression test and the true station record. The result of SRT-PS is expressed as a confidence probability ranging from 0 to 1, where a value closer to 1 indicates less uncertainty. The potential of SRT-PS to estimate quantitatively the uncertainty in temperature observations was demonstrated using an annual temperature dataset for China for the period 1971–2000 with seeded errors. SRT-PS was also applied to assess a real dataset, and was compared with two traditional quality control approaches: biweight mean and biweight standard deviation and SRT. The study provides a new approach to assess quantitatively the uncertainty in temperature observations at meteorological stations.

1. Introduction

Surface air temperature is one of the most important climate elements affecting the biosphere and human activity (Vinnikov et al. 1990). It is central to a multitude of assessments and models for a wide variety of geographic processes (Feng et al. 2004; Jeffrey et al. 2001). There are often unknown errors in the long-term temperature records of meteorological stations, which relate mainly to measurement, homogenization adjustment, calculation, and reporting (Brohan et al. 2006; Yan and Jones 2008); however, many users regard the records as

accurate and ignore any errors. Such errors may corrupt the output of a quantitative model. This problem can be minimized by the rigorous quality control (QC) of climate data (Daly et al. 2004; Durre et al. 2010; Lawrimore et al. 2011).

Traditional QC methods can be divided into two principal categories: temporal methods and spatial methods (Eischeid et al. 1995). For a target station, the temporal QC method is based on the premise that an individual observation should be statistically similar to observations made for the same time (e.g., day or month) in other years. QC is then performed using the sample distribution for each station over a designated period, and possible invalid records are flagged according to a threshold determined as a multiple of the interquartile range calculated for each station in a period or using other statistical methods (Eischeid et al. 1995; Feng et al. 2004). Temporal QC is easy to implement and has advantages, particularly when there are insufficient neighbors for spatial QC. However, an outlier observation may be due to real changes in the climate (Peterson et al. 1998) and the threshold should be determined from

* Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JAMC-D-13-0179.s1>.

Corresponding author address: Jin-Feng Wang, State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.
E-mail: wangjf@igsrr.ac.cn

experience or exploratory data analysis. The method fails if there are missing observations for a long period in the target-station record. In contrast, spatial QC methods use nearby observations with adequate data to estimate the value at a target station. Threshold limits are established according to the statistical distribution of differences between the observed and predicted (i.e., estimated) values, and a record with a difference larger than a designated threshold is deemed to include possible errors (Eischeid et al. 1995). Various interpolation methods have been developed to estimate the observation at a target station; these methods differ in their concept and mathematical formulation (Burrough and McDonnell 1998). The choice of method is based on a number of factors, such as the geographical characteristics of the station location and the spatial distribution of stations (Eischeid et al. 1995). Regression-based, geostatistical, and inverse-distance weighting methods are used widely to estimate climate datasets. In regression-based approaches, records taken from surrounding stations or external information as independent variables are employed to develop a regression model (Daly 2006; Stahl et al. 2006). Regression models that are more sophisticated have also been introduced; for example, geographically weighted regression (Fotheringham et al. 2002), spatial-lag and spatial-error models (Fischer and Wang 2011), the Parameter-Elevation Regressions on Independent Slopes Model (Daly et al. 2002) and the spatial regression test (SRT) (Hubbard and You 2005; Hubbard et al. 2007, 2005; You and Hubbard 2006; You et al. 2008, 2007). Inverse-distance weighting is the method used most commonly in climate data interpolation (Di Piazza et al. 2011); a weighted average is calculated by taking the inverse distance between the target and neighboring stations. Geostatistics (Goovaerts 1997; Isaaks and Srivastava 1989; Olea 1999) have also been applied widely to climate data interpolation (Boer et al. 2001; Hudson and Wackernagel 1994; Jeffrey et al. 2001), and they provide the best linear unbiased estimates (Burrough et al. 1998). The point estimation model of the Biased Sentinel Hospitals-Based Area Disease Estimation (P-BSHADE) is a newly developed technique (Xu et al. 2013) that uses a weighted summation of station observations to derive best linear unbiased estimates for the target station. In contrast to geostatistics, P-BSHADE considers spatial autocorrelation and heterogeneity, and maximizes an objective function for estimation.

The above traditional QC methods can generally be described as categorical and deterministic in that a series of quality checks are employed to determine whether an observation is valid. Typically, the outcomes of these checks are flagged valid or invalid (Daly et al. 2004). However, nowadays there are various quantitative models that depend on quantitative climate data as input (Feng

et al. 2004; Patz et al. 2005), and it is necessary to assess quantitatively the uncertainty in that data. In addition, different models have different tolerances to the uncertainty of input data. For example, some models require all input data to be correct (Wang et al. 2009), whereas other models can employ input data with some uncertainty (Christakos and Serre 2000; Christakos et al. 2001). Therefore, the quantitative assessment of input data is more meaningful than categorical validity (Daly et al. 2004). To respond to this demand, this paper proposes a probabilistic spatiotemporal approach based on a spatial regression test (SRT-PS). The method uses a continuous quantitative probability to represent the uncertainty level in each temperature observation at a meteorological station.

2. Method and data

a. SRT

The SRT (Hubbard et al. 2005) uses data recorded by neighboring stations to estimate the record at a target station; the value for the target station is calculated using the weighted average of the estimates. In this process, weights are determined according to the strength of a statistical relationship, such as the root-mean-square error (RMSE), between the target station and each neighboring station. Estimation by employing the SRT is described briefly in the following.

First, for each neighboring station, a regression-based estimate $x_i = a + by_i$ is obtained for each pair x_i and y_i , where x is the record for the target station, y_i is the record for the i th neighboring station ($i = 1, \dots, n$), and a and b are regression coefficients. Then, for an observed y_i , estimates of x_i are calculated. A weighted estimate $\hat{x}$ of these n estimates of x is obtained using the RMSE of the n regression estimates:

$$\hat{x} = \pm \left(\sum_{i=1}^N \pm \frac{x_i^2}{se_i^2} \right)^{1/2} \bigg/ \left(\sum_{i=1}^N \frac{1}{se_i^2} \right)^{1/2}, \quad (1)$$

where N is the number of stations used in the estimate, se_i is the RMSE for neighboring station y_i , and $\hat{x}$ and x_i must have the correct sign (You and Hubbard 2006).

b. Uncertainty estimation employing the SRT

SRT-PS evaluates the uncertainty in the target-station observation in three steps. (i) For each target-station observation, an estimate is obtained in an SRT and a residual (i.e., the difference between the estimate and the actual observation) is obtained. (ii) The expectation and standard deviation of the residuals are estimated. (iii) The confidence probability (cp) for each record is

FIG. 1. Probability distribution of residuals for neighbors of the target station.

calculated. The p value of the residual for each record is estimated from the probability distribution determined by the residuals of the spatiotemporal neighbors. In this study, the cp is expressed by the p value, which reflects the uncertainty in the observation:

$$cp = p = 2P\left[z > \frac{|O(e) - E(e)|}{S(e)}\right], \quad (2)$$

$cp \in [0, 1]$, where a value closer to 1 indicates less uncertainty, e is the residual for each record estimated by SRT, $E(e)$ and $S(e)$ are the expectation and standard deviation for the probability distribution respectively, $O(e)$ is the observed residual of the target station, z is the z value in standard normal distribution, and p is the p value for $O(e)$ in the probability distribution (Fig. 1). The assumption of SRT-PS is that if there is a large residual difference between the estimated value and actual observation, there might be large uncertainty in the station record. The reliability of SRT has been demonstrated in several studies (Hubbard and You 2005; Hubbard et al. 2007, 2005).

For each target station, the m spatial neighbors and n temporal neighbors are used; therefore, the total number of neighboring stations is $(m + 1) \times n - 1$ (Fig. 2). In this study, m is 10 and n is 9 and the probability distribution determined by neighboring residuals is assumed to satisfy a normal distribution.

c. Validation

To document the performance of the method in a controlled situation, seeded errors were introduced to the datasets of annual station observations in China for the period 1971–2000. The potential of the method to quantify these errors was recorded. The period of 1971–2000 was selected because it has the highest data quality of the historical period and it was suitable for the identification of seeded errors. The magnitude of the seeded

FIG. 2. Target station and its selected spatiotemporal neighbors.

error was determined randomly. A random number f was selected using a random-number generator operating on a uniform distribution with a mean of zero and range of $[-3.5, 3.5]$. The selected range ensures that the tests include cases that are close to the extremes of the period (Hubbard et al. 2007, 2005). Here, f also represents the standardized anomalies of the seeded errors. This number was then multiplied by the standard deviation S of the variable in the period 1971–2000 to obtain the error magnitude E for the randomly selected observation x :

$$E_x = S_x f. \quad (3)$$

Thus, the expected distribution of the error magnitude has a mean of zero and a range equal to 3.5 times the standard deviation of the variable S_x . Errors were seeded into temperature records by 1) randomly drawing a year from 1971–2000, 2) randomly drawing a station from all

FIG. 3. Distribution of stations (dots) and topography (colored shading, m).

stations in the selected year, 3) generating a random number f , and 4) calculating an error magnitude E .

d. Data

An annual climate dataset for the period 1971–2000 was used to validate the method. This dataset was constructed by the Chinese National Meteorological Center, and was quality controlled and homogenized by Li et al. (2004, 2010). However, this does not mean that the records in the dataset were true values, and small errors might not have been identified and adjusted in the homogenization. The number of stations in each year for the period is 582. During the study period, stations were distributed evenly across eastern China, whereas the density of stations was sparse in western and northern parts of the country (Fig. 3).

3. Test of SRT-PS

In the test of SRT-PS, a station was selected randomly from all stations for the period 1971–2000 and a random error was seeded in the record of the selected station using the method described above. The cp value was then estimated. By repeating the above process 1000

times, 1000 cp values were obtained for the 1000 randomly selected stations.

The relationship between the absolute seeded error and cp is presented in Fig. 4. The figure shows that cp decreased with an increase in absolute seeded error, and that most cp values were 0 when the absolute seeded error was >0.3 . For absolute seeded errors <0.3 , cp decreased gradually, the coefficient of correlation between the absolute seeded error and cp was -0.90 , and the relationship was statistically significant (p value of 0). In the test, the uncertainty of all station values with seeded errors was assessed. A cp value near 1 indicates low uncertainty, whereas a cp value near 0 indicates high uncertainty; this is distinctly different from traditional QC methods.

In some cases, we would be interested in station records with low cp values, and in some applications, a predefined threshold (e.g., $cp < 0.05$) may be used to determine the values with high uncertainty; the threshold is usually different in different applications. In the study, as an example, the relationship between the average absolute seeded error in each interval and cp was analyzed for the threshold of $cp < 0.05$. When the absolute seeded error was between 0.0 and 0.1°C, no

FIG. 4. Scatterplots of absolute seeded error vs cp.

stations were identified and the proportion increased for higher absolute seeded errors; 84.62% of stations were identified when the absolute seeded error was between 0.4 and 0.5°C, and almost all stations were identified when the absolute seeded error was $>0.6^{\circ}\text{C}$.

4. Uncertainty estimation for a real dataset

We applied the SRT-PS to a real dataset and compared it with two alternative approaches: the biweight mean and biweight standard deviation method (abbreviated as biweight below) (Lanzante 1996) and SRT (You and Hubbard 2006; You et al. 2008). The details of the two methods are described in the appendix.

Figure 5 presents the frequency of the cp values estimated by the SRT-PS method. Generally, most of the stations have cp values higher than 0.5, which means that most of the temperature records have higher confidence probability. For the records with high or low confidence probability, the SRT-PS estimation shows that there are 5.29% of records that have cp values smaller than 0.05, and 5.1% of records that have cp values smaller than 0.95.

Figure 6a presents the results calculated by SRT-PS with cp values smaller than three standard deviations of their corresponding probability distribution, and 78 stations were found to be abnormal. These stations were mainly around the eastern edge of the Qinghai-Tibet Plateau, encompassing the northwest, northeast, and southwest of China. Figure 6b shows the results calculated by SRT with temperature records outside the confidence limit determined by 3 times the weighted

RMSE [see Eq. (A5)]; 31 stations were found to be abnormal, and these stations were mainly in the interior of the Qinghai-Tibet Plateau. The temporal distribution of these records detected by the above two methods were relatively even during the 30 yr. Figure 6c shows the results calculated by biweight with temperature records greater than the critical threshold value of three biweight standard deviations [see Eq. (A3)]. Station distributions detected by biweight were mainly in

FIG. 5. Frequency of cp values estimated by SRT-PS by a real dataset.

FIG. 6. Spatial distribution of stations (dots) with (a) large uncertainty estimated by SRT-PS, (b) anomaly detected by SRT, and (c) biweight; and topography (gray shading, m).

northeastern and southern China; the total number is 163, and the temporal distribution of these records is concentrated in the early and last few years (Fig. 6c).

5. Discussion and conclusions

An SRT-based probabilistic spatiotemporal approach (SRT-PS) was proposed for estimating the uncertainty in temperature observations at meteorological stations. The assumption of SRT-PS is that if there is a large residual difference between the estimated and actually recorded values, there might be large uncertainty in the station record. Thus, making use of the probability distribution determined by the residuals of spatiotemporal neighbors of the target station, we obtained a continuous quantitative probability representing the uncertainty in each temperature observation at the target station. A validation study, using an annual temperature dataset for China with seeded errors, demonstrated the effective potential of SRT-PS in estimating quantitatively the uncertainty in temperature observations.

Traditional QC methods can generally be classified as categorical and deterministic in that a threshold range is determined using neighboring stations in the spatial, temporal, or spatiotemporal domain. QC requires that an observation must lie within the threshold range if it is to be considered valid (Hubbard et al. 2005). The outcomes of such checks are typically flagged yes or no. However, such qualitative QC is unsuitable for real datasets. Errors in station observations have many sources, such as homogenization adjustment, calculation, and reporting. Many of these errors cannot be validated because of missing metadata, and only some extreme errors can be detected by traditional methods (Brohan et al. 2006). In theory, it is possible for each station observation to contain error, and it is simply the case that the errors differ in magnitude. Therefore, it is more reasonable to estimate quantitatively the uncertainty in station observations.

There are also some limits in the results of the present study. The cp value of each station is estimated in the SRT of the observation of a station and those of the neighboring stations, and the reliability of the relationship between the stations as well as the model itself affects the accuracy of the cp value, which is not calibrated to represent the probability of uncertainty. There are five main sources of uncertainty in the model. 1) There are random fluctuations in station temperature (Brohan et al. 2006); for example, cold fronts may have different effects on measurements made at nearby stations because of differences in topography and location, thus destabilizing the relationship between stations. 2) Station observations themselves contain errors relating to

changes in the geographic environment of the station location, homogenization adjustment, calculation, and reporting (Pielke et al. 2007). These errors may make the SRT unstable. The analysis of error propagation from neighboring stations is presented in the supplemental material. 3) A sparse distribution of stations introduces uncertainty (Pielke et al. 2007). For example, before the 1950s, the spatial distribution of stations in China was very sparse with most stations located in the east of the country. The number of stations increased sharply until stabilizing around 1960. However, the distribution is still relatively sparse in western and northern parts of the country. Thus, areas with a low density of stations introduce uncertainty in the linear regression model. For example, a sensitivity of the abnormality detection to the number of neighboring stations used in SRT-PS has been analyzed. When using spatial-only or temporal-only neighboring stations, the coefficients of correlation between the absolute seeded error and cp were -0.82 and -0.78 , respectively (absolute seeded error < 0.05). This is obviously smaller than the correlation using spatiotemporal neighboring stations, as presented in the section of the test of SRT-PS. In addition, in the two scenarios, there are only 44.8% and 41.1% of cases that satisfy the assumption of a normal distribution, which is also lower than the scenario of using spatiotemporal neighboring stations (84%). 4) There is uncertainty relating to the model itself. The relationship between the station temperature and the temperatures recorded by the neighboring stations cannot be expressed completely by the SRT. For example, there are 10 assumptions for linear regression (Gujarati 2003), each of which has been checked by the real dataset used in the study, and it was found that there were 28% of cases that did not fit the assumption of no autocorrelation between the disturbances, in which case the estimated parameters are still unbiased, but they are no longer efficient. Furthermore, if records are missing for a long period, the SRT fails. 5) The cp value of the target station is estimated from the normal distribution deduced by the probability distribution of residuals for the neighboring stations. The true probability distribution is complex and therefore there might be uncertainty in the cp value. For example, there are 16% of the distributions that did not satisfy the normal assumption within the 95% confidence level.

There were 31 stations detected by the SRT method, of which 24 stations can be found in the results of the SRT-PS (Fig. 6), because SRT-PS is an extension of SRT and the residuals, as the input of SRT-PS, were calculated by SRT. Meanwhile, stations detected by biweight were concentrated in the early and later years, because this temporal QC method may misjudge the station values at the beginning or the end of the series because

there is no spatial comparison in the method. In addition, in the study, the stations with high uncertainty estimated by SRT-PS were compared with biweight and SRT. However, in the strict sense, the thresholds of the three methods do not have the same meanings because the thresholds were calculated by different methods. For example, the threshold in the SRT-PS was calculated by three standard deviations of the corresponding probability distribution of residuals from neighboring stations, the threshold of SRT was determined by 3 times the weighted RMSE [Eq. (A5)], whereas in biweight, the threshold was deduced as three biweight standard deviations [Eq. (A3)].

In future work, other sophisticated models will be selected to improve the uncertainty estimation. For example, Biased Sample Hospital-Based Area Disease Estimation (B-SHADE; Wang et al. 2011) and Mean of Surface with Nonhomogeneity (MSN; Wang et al. 2009) can be used to interpolate missing station records, while accounting for geographic spatial autocorrelation and nonhomogeneity, and maximizing an objective function for the best linear unbiased estimates of the target station. In addition, in the SRT-PS method, temporal neighbors are treated the same as spatial neighbors. In future studies, we will consider giving neighboring stations different weights according to the analysis of the spatiotemporal variation of temperature.

Acknowledgments. This study was supported by CAS (XDA05090102), NSFC (41023010; 41271404), and MOST (2012CB955503; 2012ZX10004-201; 2011AA120305) grants.

APPENDIX

Description of Quality Control Approaches Used in the Study

a. Biweight mean and biweight standard deviation

Biweight is a method by which to check for temporal outliers with the advantages of being resistant, robust, nonparametric, and usually much less affected by the outliers. The estimate is calculated by averaging samples that are weighted by their distance from the center of the distribution; the further the distance, the smaller the weighting received (Lanzante 1996). The weight w_k is calculated as follows:

$$w_k = (X_k - M) \times (c \times \text{MAD})^{-1}, \quad (\text{A1})$$

where X_k is the temperature record in the k th station; M and MAD are the median and the median of the absolute values of the deviations of the station temperature

values from the median; c is a constant, which is recommended to be set between 6 and 9 (Lanzante 1996); and in this study $c = 7.5$. In addition, w_k was set to be 0 if $|w_k| \geq 1$.

The biweight estimate of the mean $\bar{X}$ is quantified as

$$\bar{X} = M + \left[\sum_{k=1}^n (X_k - M)(1 - w_k^2)^2 \right] \left[\sum_{k=1}^n (1 - w_k^2)^2 \right]^{-1}, \quad (\text{A2})$$

and the biweight estimate of the standard deviation S is expressed as

$$S = n \left[\sum_{k=1}^n (X_k - M)(1 - w_k^2)^4 \right]^{0.5} \times \left[\sum_{k=1}^n (1 - w_k^2)(1 - 5w_k^2) \right]^{-1}. \quad (\text{A3})$$

The values of $\bar{X}$ and S are used to determine the Z score of the temperature value at a target station by

$$Z = |X_o - \bar{X}| \times S^{-1} \quad (\text{A4})$$

where X_o is the target value; Z -score values greater than a certain critical threshold value will be regarded as outliers. In the study, the criterion $Z \geq 3$ was used.

b. Quality control using SRT

The quality of the temperature value at a target station can be assessed by SRT [see Eq. (1)], in which a confidence interval is employed. This is defined as

$$\hat{x} - f \times \text{se}' \leq x \leq \hat{x} + f \times \text{se}', \quad (\text{A5})$$

where $\hat{x}$ has the same meaning as in Eq. (1) in the main text, se' is the weighted standard error of the estimate $\text{se}' = N^{-1} \sum_{i=1}^N \text{se}_i^{-2}$, f is a constant used to define the desired confidence limit, and in the present study $f = 3$. If the temperature at the target station causes the relation to be true, then the corresponding value passes the test.

REFERENCES

- Boer, E. P. J., K. M. de Beurs, and A. D. Hartkamp, 2001: Kriging and thin plate splines for mapping climate variables. *Int. J. Appl. Earth. Obs. Geoinf.*, **3**, 146–154, doi:10.1016/S0303-2434(01)85006-6.
- Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.*, **111**, D12106, doi:10.1029/2005JD006548.
- Burrough, P. A., and R. A. McDonnell, 1998: *Principles of Geographical Information Systems*. 2nd ed. Oxford University Press, 352 pp.
- Christakos, G., and M. L. Serre, 2000: BME analysis of spatiotemporal particulate matter distributions in North Carolina. *Atmos. Environ.*, **34**, 3393–3406, doi:10.1016/S1352-2310(00)00080-7.
- , —, and J. L. Kovitz, 2001: BME representation of particulate matter distributions in the state of California on the basis of uncertain measurements. *J. Geophys. Res.*, **106**, 9717–9731, doi:10.1029/2000JD900780.
- Daly, C., 2006: Guidelines for assessing the suitability of spatial climate data sets. *Int. J. Climatol.*, **26**, 707–721, doi:10.1002/joc.1322.
- , W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris, 2002: A knowledge-based approach to the statistical mapping of climate. *Climate Res.*, **22**, 99–113, doi:10.3354/cr022099.
- , —, M. Doggett, J. Smith, and G. Taylor, 2004: A probabilistic-spatial approach to the quality control of climate observations. Preprints, *14th Conf. on Applied Climatology*, Seattle, WA, Amer. Meteor. Soc., 7.3. [Available online at <https://ams.confex.com/ams/84Annual/webprogram/Paper71411.html>.]
- Di Piazza, A., F. Lo Conti, L. V. Noto, F. Viola, and G. La Loggia, 2011: Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. *Int. J. Appl. Earth. Obs. Geoinf.*, **13**, 396–408, doi:10.1016/j.jag.2011.01.005.
- Durre, I., M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, 2010: Comprehensive automated quality assurance of daily surface observations. *J. Appl. Meteor. Climatol.*, **49**, 1615–1633, doi:10.1175/2010JAMC2375.1.
- Eischeid, J. K., C. B. Baker, T. R. Karl, and H. F. Diaz, 1995: The quality control of long-term climatological data using objective data analysis. *J. Appl. Meteor.*, **34**, 2787–2795, doi:10.1175/1520-0450(1995)034<2787:TQCOLT>2.0.CO;2.
- Feng, S., Q. Hu, and W. Qian, 2004: Quality control of daily meteorological data in China, 1951–2000: A new dataset. *Int. J. Climatol.*, **24**, 853–870, doi:10.1002/joc.1047.
- Fischer, M. M., and J.-F. Wang, 2011: *Spatial Data Analysis: Models, Methods and Techniques*. Springer, 80 pp.
- Fotheringham, A. S., C. Brunsdon, and M. Charlton, 2002: *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons, Inc., 284 pp.
- Goovaerts, P., 1997: *Geostatistics for Natural Resources Evaluation*. Oxford University Press, 496 pp.
- Gujarati, D. N., 2003: *Basic Econometrics*. 4th ed. McGraw-Hill, 1002 pp.
- Hubbard, K. G., and J. You, 2005: Sensitivity analysis of quality assurance using the spatial regression approach—A case study of the maximum/minimum air temperature. *J. Atmos. Oceanic Technol.*, **22**, 1520–1530, doi:10.1175/JTECH1790.1.
- , S. Goddard, W. D. Sorensen, N. Wells, and T. T. Osugi, 2005: Performance of quality assurance procedures for an applied climate information system. *J. Atmos. Oceanic Technol.*, **22**, 105–112, doi:10.1175/JTECH-1657.1.
- , N. B. Guttman, J. You, and Z. R. Chen, 2007: An improved QC process for temperature in the daily cooperative weather observations. *J. Atmos. Oceanic Technol.*, **24**, 206–213, doi:10.1175/JTECH1963.1.
- Hudson, G., and H. Wackernagel, 1994: Mapping temperature using kriging with external drift: Theory and an example from Scotland. *Int. J. Climatol.*, **14**, 77–91, doi:10.1002/joc.3370140107.
- Isaaks, E. H., and R. M. Srivastava, 1989: *An Introduction to Applied Geostatistics*. Oxford University Press, 561 pp.
- Jeffrey, S. J., J. O. Carter, K. B. Moodie, and A. R. Beswick, 2001: Using spatial interpolation to construct a comprehensive archive

- of Australian climate data. *Environ. Modell Software*, **16**, 309–330, doi:[10.1016/S1364-8152\(01\)00008-1](https://doi.org/10.1016/S1364-8152(01)00008-1).
- 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. *Int. J. Climatol.*, **16**, 1197–1226, doi:[10.1002/\(SICI\)1097-0088\(199611\)16:11<1197::AID-JOC89>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-0088(199611)16:11<1197::AID-JOC89>3.0.CO;2-L).
- Lawrimore, J. H., M. J. Menne, B. E. Gleason, C. N. Williams, D. B. Wuertz, R. S. Vose, and J. Rennie, 2011: An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *J. Geophys. Res.*, **116**, D19121, doi:[10.1029/2011JD016187](https://doi.org/10.1029/2011JD016187).
- Li, Q., H. Zhang, X. Liu, and J. Huang, 2004: Urban heat island effect on annual mean temperature during the last 50 years in China. *Theor. Appl. Climatol.*, **79**, 165–174, doi:[10.1007/s00704-004-0065-4](https://doi.org/10.1007/s00704-004-0065-4).
- , W. Dong, W. Li, X. Gao, P. Jones, J. Kennedy, and D. Parker, 2010: Assessment of the uncertainties in temperature change in China during the last century. *Chin. Sci. Bull.*, **55**, 1974–1982, doi:[10.1007/s11434-010-3209-1](https://doi.org/10.1007/s11434-010-3209-1).
- Olea, R. A., 1999: *Geostatistics for Engineers and Earth Scientists*. Kluwer Academic, 303 pp.
- Patz, J. A., D. Campbell-Lendrum, T. Holloway, and J. A. Foley, 2005: Impact of regional climate change on human health. *Nature*, **438**, 310–317, doi:[10.1038/nature04188](https://doi.org/10.1038/nature04188).
- Peterson, T. C., and Coauthors, 1998: Homogeneity adjustments of *in situ* atmospheric climate data: A review. *Int. J. Climatol.*, **18**, 1493–1517, doi:[10.1002/\(SICI\)1097-0088\(199811\)18:13<1493::AID-JOC329>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1097-0088(199811)18:13<1493::AID-JOC329>3.0.CO;2-T).
- Pielke, R. A., Sr., and Coauthors, 2007: Unresolved issues with the assessment of multidecadal global land surface temperature trends. *J. Geophys. Res.*, **112**, D24S08, doi:[10.1029/2006JD008229](https://doi.org/10.1029/2006JD008229).
- Stahl, K., R. D. Moore, J. A. Floyer, M. G. Asplin, and I. G. McKendry, 2006: Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. *Agric. For. Meteorol.*, **139**, 224–236, doi:[10.1016/j.agrformet.2006.07.004](https://doi.org/10.1016/j.agrformet.2006.07.004).
- Vinnikov, K. Ya, P. Ya Groisman, and K. M. Lugina, 1990: Empirical data on contemporary global climate changes (temperature and precipitation). *J. Climate*, **3**, 662–677, doi:[10.1175/1520-0442\(1990\)003<0662:EDOCGC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1990)003<0662:EDOCGC>2.0.CO;2).
- Wang, J.-F., G. Christakos, and M.-G. Hu, 2009: Modeling spatial means of surfaces with stratified nonhomogeneity. *IEEE Trans. Geosci. Remote Sens.*, **47**, 4167–4174, doi:[10.1109/TGRS.2009.2023326](https://doi.org/10.1109/TGRS.2009.2023326).
- , and Coauthors, 2011: Area disease estimation based on sentinel hospital records. *PLoS ONE*, **6**, e23428, doi:[10.1371/journal.pone.0023428](https://doi.org/10.1371/journal.pone.0023428).
- Xu, C.-D., J.-F. Wang, M.-G. Hu, and Q.-X. Li, 2013: Interpolation of missing temperature data at meteorological stations using P-BSHADE. *J. Climate*, **26**, 7452–7463, doi:[10.1175/JCLI-D-12-00633.1](https://doi.org/10.1175/JCLI-D-12-00633.1).
- Yan, Z. W., and P. D. Jones, 2008: Detecting inhomogeneity in daily climate series using wavelet analysis. *Adv. Atmos. Sci.*, **25**, 157–163, doi:[10.1007/s00376-008-0157-7](https://doi.org/10.1007/s00376-008-0157-7).
- You, J. S., and K. G. Hubbard, 2006: Quality control of weather data during extreme events. *J. Atmos. Oceanic Technol.*, **23**, 184–197, doi:[10.1175/JTECH1851.1](https://doi.org/10.1175/JTECH1851.1).
- , S. Nadarajah, and K. E. Kunkel, 2007: Performance of quality assurance procedures on daily precipitation. *J. Atmos. Oceanic Technol.*, **24**, 821–834, doi:[10.1175/JTECH2002.1](https://doi.org/10.1175/JTECH2002.1).
- , K. G. Hubbard, and S. Goddard, 2008: Comparison of methods for spatially estimating station temperatures in a quality control system. *Int. J. Climatol.*, **28**, 777–787, doi:[10.1002/joc.1571](https://doi.org/10.1002/joc.1571).